GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 16:58:15 ; Search time 46.08 Seconds
(without alignments)
5189.123 Million cell updates/sec

Title: US-09-037-460-1
Perfect score: 1271
Sequence: 1 CTGCTTCCCACCAGCAAAGA.....GTTNCCTNTNTCTNAAGGAC 1271

Scoring table: OLIGO_NUC

Searched: 240622 seqs, 94065609 residues

word = 30

Database : N_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1220	96.0	1271	1	T34991	Vascular IBP-like

ALIGNMENTS

RESULT 1
T34991
ID T34991 standard; cDNA; 1271 BP.
AC T34991;
DT 06-NOV-1996 (first entry)
DE Vascular IBP-like growth factor cDNA.
KW Vascular IBF-like growth factor; VIGF;
KW insulin-like growth factor binding protein; agonist; antagonist;
KW muscle wastage; osteoporosis; implant fixation; wound healing;
KW therapy; diagnosis; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 58. .612
FT /*tag= a
FT signal_peptide 58. .120
FT /*tag= b
FT mat_peptide 121. .609
FT /*tag= c
FT misc_difference 653
FT /*tag= d
FT /note= "base 653 is given as y in the
FT specification"
FT misc_difference 1096
FT /*tag= e
FT /note= "base 1096 is given as r in the
FT specification"
FT misc_difference 1212
FT /*tag= f
FT /note= "base 1212 is given as n in the
FT specification"
FT misc_difference 1225
FT /*tag= g
FT /note= "base 1225 is given as y in the
FT specification"
FT misc_difference 1255
FT /*tag= h
FT /note= "base 1255 is given as n in the
FT specification"
FT misc_difference 1259
FT /*tag= i
FT /note= "base 1259 is given as n in the
FT specification"
FT misc_difference 1261
FT /*tag= j
FT /note= "base 1261 is given as n in the
FT specification"
FT misc_difference 1265
FT /*tag= k
FT /note= "base 1265 is given as n in the
FT specification"
PN WO9617931-A1.
PD 13-JUN-1996.
PF 09-DEC-1994; U14388.
PR 09-DEC-1994; WO-U14388.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Hastings GA, Rosen CA;
DR WPI; 96-287176/29.
DR P-PSDB; R98994.
PT Human vascular insulin-like growth factor binding protein-like
PT growth factor, and its nucleic acid sequence and (ant)agonists -
PT used, e.g. to treat muscle wasting diseases or aid implant fixation,
PT or limit excess connective tissue prodn. during wound healing.
PS Claim 1; Page 42-43; 61pp; English.
CC A cDNA clone (T34991) codes for human vascular insulin-like
CC growth factor binding protein-like growth factor (R98994),
CC or VIGF, a protein structurally related to the IBP and CCN
CC families. It was discovered in a cDNA library derived from human
CC umbilical vein endothelial cells. The cDNA can be used for
CC prodn. of recombinant VIGF, e.g. by expression in E. coli, CHO
CC or insect host cells. It is also useful therapeutically e.g.
CC to treat muscle wasting diseases or osteoporosis, or to design
CC probes for the detection of diseases associated with under- or
CC over-expression of VIGF.
SQ Sequence 1271 BP; 360 A; 246 C; 337 G; 320 T;

Query Match 96.0%; Score 1220; DB 1; Length 1271;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60

Qy      61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120

Qy     121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180

Qy     181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240

Qy     241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGGCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300

Qy     301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360

Qy     361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420

Qy     421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480

Qy    481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540

Qy    541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600

Qy    601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAA 660

Qy    661 CACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAA 720

Qy    721 GACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780

Qy    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840

Qy    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900

Qy    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960

Qy    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020

Qy   1021 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGATGGGGGGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080

Qy   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140

Qy   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200

Qy   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260

Qy   1261 NTCTNAAGGAC 1271
          ||||||||||||
Db   1261 NTCTNAAGGAC 1271
```

Search completed: May 3, 1999, 17:14:00
Job time: 945 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 17:42:14 ; Search time 358.71 Seconds
(without alignments)
5404.963 Million cell updates/sec

Title: US-09-037-460-1
Perfect score: 1271
Sequence: 1 CTGCTTCCCACCAGCAAAGA.....GTTNCCTNTNTCTNAAGGAC 1271

Scoring table: OLIGO_NUC

Searched: 2002476 seqs, 762712212 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: gb_est1:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: gb_est6:*
16: gb_est7:*
17: gb_est8:*
18: gb_est9:*
19: gb_est10:*
20: gb_est11:*
21: gb_est12:*
22: gb_est13:*
23: gb_est22:*
24: gb_est14:*
25: gb_est15:*
26: gb_est16:*
27: gb_est17:*
28: gb_est18:*
29: gb_est19:*
30: gb_est20:*
31: gb_est21:*
32: em_est10:*
33: em_est11:*
34: em_est12:*
35: em_est13:*
36: em_est14:*
37: em_est15:*
38: em_est16:*
39: em_est17:*
40: em_est18:*
41: em_est19:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	209	16.4	425	14	W46667	W46667 zc33e02.r1
2	48	3.8	351	17	AA296697	AA296697 EST112419

ALIGNMENTS

RESULT 1
W46667
LOCUS W46667 425 bp mRNA EST 11-OCT-1996
DEFINITION zc33e02.r1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone 324122 5', mRNA sequence.
ACCESSION W46667
NID g1331295
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 425)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1322 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 395.
FEATURES Location/Qualifiers
source 1. .425
/organism="Homo sapiens"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI; TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="324122"
/clone_lib="Soares senescent fibroblasts NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
mRNA <1. .>425
BASE COUNT 144 a 50 c 102 g 126 t 3 others
ORIGIN

Query Match 16.4%; Score 209; DB 14; Length 425;
Best Local Similarity 100.0%; Pred. No. 6e-93;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    812 GTTAAATATTCGAATGCATGTAGATTTGTTAAATGTGTGTGTATAGTAACACTGAAGAAC 871
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTTAAATATTCGAATGCATGTAGATTTGTTAAATGTGTGTGTATAGTAACACTGAAGAAC 60

Qy    872 TAAAAATGCAATTTAGGTAATCTTACATGGAGACAGGTCAACCAAAGAGGGAGCTAGGCA 931
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TAAAAATGCAATTTAGGTAATCTTACATGGAGACAGGTCAACCAAAGAGGGAGCTAGGCA 120

Qy    932 AAGCTGAAGACCGCAGTGAGTCAAATTAGTTCTTTGACTTTC
```

```
                                                                    TAATGTTGGG 991
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGCTGAAGACCGCAGTGAGTCAAATTAGTTCTTTGACTTTGATGTACATTAATGTTGGG 180

Qy    992 ATATGGAATGAAGACTTAAGAGCAGGAGA 1020
          ||||||||||||||||||||||||||||
Db    181 ATATGGAATGAAGACTTAAGAGCAGGAGA 209


RESULT   2
AA296697
LOCUS       AA296697      351 bp    mRNA            EST       18-APR-1997
DEFINITION  EST112419 Aorta endothelial cells, TNF alpha-treated Homo sapiens
            cDNA 5' end similar to endothelial cell-specific secretory
            molecule, mRNA sequence.
ACCESSION   AA296697
NID         g1949189
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC154673
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .351
                      /organism="Homo sapiens"
                      /note="Organ: aorta; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):115698"
                      /db_xref="taxon:9606"
                      /clone_lib="Aorta endothelial cells, TNF alpha-treated"
                      /cell_type="endothelial cell"
                      /dev_stage="adult"
     mRNA             <1. .>351
BASE COUNT      118 a     57 c     72 g    102 t      2 others
ORIGIN


  Query Match             3.8%;  Score 48;  DB 17;  Length 351;
  Best Local Similarity   100.0%;  Pred. No. 2e-13;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1123 GGGAAAAGTCAAAATAAAACAACCAGAAAACCCCTGAAGGAAGTAAGA 1170
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGGAAAAGTCAAAATAAAACAACCAGAAAACCCCTGAAGGAAGTAAGA 48


Search completed: May  3, 1999, 17:53:24
Job time: 670 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 16:05:31 ; Search time 12.08 Seconds
(without alignments)
408.826 Million cell updates/sec

Title: US-09-037-460-2
Perfect score: 184
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....EVVKENAAGSPVMRKWLNPR 184

Scoring table: OLIGO

Searched: 74019 seqs, 26840295 residues

Word = 30

Database : SwissProt_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

no matches

ALIGNMENTS

Search completed: May 3, 1999, 16:57:31
Job time: 3120 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 20:41:34 ; Search time 16.01 Seconds
(without alignments)
430.525 Million cell updates/sec

Title: US-09-037-460-2
Perfect score: 184
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....EVVKENAAGSPVMRKWLNPR 184

Scoring table: OLIGO

Searched: 116738 seqs, 37460341 residues word = 30

Database : PIR_58:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

no matches

ALIGNMENTS

Search completed: May 3, 1999, 21:34:32
Job time: 3178 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 17:54:24 ; Search time 47.06 Seconds
(without alignments)
17.585 Million cell updates/sec

Title: US-09-037-460-2_COPY_55_69
Perfect score: 78
Sequence: 1 RVCAAGRGETCYRTV 15

Scoring table: PAM150

Searched: 180763 seqs, 55169189 residues

Database : SPTREMBL_8:*
1: sp_fungi:*
2: sp_human:*
3: sp_invertebrate:*
4: sp_mammal:*
5: sp_mhc:*
6: sp_organelle:*
7: sp_phage:*
8: sp_plant:*
9: sp_bacteria:*
10: sp_rodent:*
11: sp_virus:*
12: sp_vertebrate:*
13: sp_unclassified:*
14: sp_archea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	78	100.0	184	2	Q15330	Q15330 homo sapien
2	66	84.6	184	10	P97682	P97682 rattus norv
3	41.5	53.2	313	11	O41072	O41072 paramecium
4	41	52.6	440	3	O18003	O18003 caenorhabdi
5	40	51.3	303	2	O75331	O75331 homo sapien
6	39	50.0	218	9	P94354	P94354 bacillus su
7	38	48.7	139	11	Q88593	Q88593 theiler mur
8	37	47.4	92	11	O38024	O38024 potato viru
9	39	50.0	488	8	P93694	P93694 volvox cart
10	36.5	46.8	208	3	O44335	O44335 lumbricus t
11	40	51.3	1127	3	Q94248	Q94248 caenorhabdi
12	36	46.2	179	11	P89137	P89137 simian hemo
13	38	48.7	472	2	O14522	O14522 homo sapien
14	35	44.9	137	3	Q20876	Q20876 caenorhabdi
15	40	51.3	1420	9	O52666	O52666 escherichia
16	39	50.0	1040	8	O23347	O23347 arabidopsis
17	33.5	42.9	82	12	Q98965	Q98965 naja naja (
18	34	43.6	105	9	Q56672	Q56672 vibrio chol
19	36	46.2	278	3	Q23311	Q23311 caenorhabdi
20	33	42.3	70	11	Q87620	Q87620 simian immu
21	38	48.7	716	3	O44118	O44118 drosophila
22	36	46.2	315	3	O45842	O45842 caenorhabdi
23	34	43.6	127	8	O04321	O04321 arabidopsis
24	34	43.6	130	3	O77184	O77184 plasmodium
25	34	43.6	130	12	Q90230	Q90230 ambystoma m
26	34	43.6	132	3	O77183	O77183 plasmodium
27	36	46.2	345	3	Q94573	Q94573 haemonchus
28	32.5	41.7	69	11	Q68372	Q68372 hepatitis c
29	34	43.6	145	9	O05929	O05929 prochloroth
30	34	43.6	151	11	O12669	O12669 colobus mon
31	36	46.2	387	2	O43603	O43603 homo sapien
32	36	46.2	388	9	P96847	P96847 mycobacteri
33	33.5	42.9	122	12	Q91521	Q91521 vipera ammo
34	36	46.2	395	10	Q64166	Q64166 rattus norv
35	34	43.6	157	3	O18506	O18506 strongyloid
36	36	46.2	457	14	O58493	O58493 pyrococcus
37	37	47.4	728	10	Q64007	Q64007 mus musculu
38	35.5	45.5	367	8	O22205	O22205 arabidopsis
39	35	44.9	293	9	O87716	O87716 caulobacter
40	37	47.4	748	10	Q61662	Q61662 mus musculu
41	35	44.9	301	9	Q48399	Q48399 klebsiella
42	31.5	40.4	62	12	O73855	O73855 naja sputat
43	31.5	40.4	62	12	O73856	O73856 naja sputat
44	31.5	40.4	62	12	O73857	O73857 naja sputat
45	31.5	40.4	62	12	O73859	O73859 naja sputat

ALIGNMENTS

RESULT 1
Q15330
ID Q15330 PRELIMINARY; PRT; 184 AA.
AC Q15330;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ESM-1 SECRETORY PROTEIN PRECURSOR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96355375.
RA LASSALLE P.M., MOLET S., JANIN A., VANDER-HEYDEN J.E., TAVERNIER J.,
RA FIERS W., DEVOS R.E., TONNEL A.E.B.;
RT "ESM-1 is a novel human endothelial cell-specific molecule expressed
RT in lung and regulated by cytokines.";
RL J. BIOL. CHEM. 271:20458-20464(1996).
DR EMBL; X89426; E189266; -.
DR PFAM; PF00219; IGFBP; 1.
KW SIGNAL.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 184 ESM-1 SECRETORY PROTEIN.
SQ SEQUENCE 184 AA; 20095 MW; 08D109DF CRC32;

Query Match 100.0%; Score 78; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RVCAAGRGETCYRTV 15
||||||||||||||||
Db 55 RVCAAGRGETCYRTV 69

RESULT 2
P97682
ID P97682 PRELIMINARY; PRT; 184 AA.
AC P97682;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE PG25.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY; TISSUE=PINEAL GLAND;
RA WANG X., BROWNSTEIN M.J., YOUNG W.S.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.

```
DR   EMBL; U80818; G1750198; -.
SQ   SEQUENCE   184 AA;  20075 MW;  3457564C CRC32;


  Query Match             84.6%;  Score 66;  DB 10;  Length 184;
  Best Local Similarity   86.7%;  Pred. No. 0.00057;
  Matches   13;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 RVCAAGRGETCYRTV 15
           :|||||  ||||||||
Db      55 QVCAAGPGETCYRTV 69


RESULT   3
O41072
ID   O41072      PRELIMINARY;      PRT;   313 AA.
AC   O41072;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   A590L PROTEIN.
GN   A590L.
OS   PARAMECIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98022962.
RA   LI Y., LU Z., SUN L., ROPP S., KUTISH G.F., ROCK D.L., VAN ETTEN J.L.;
RT   "Analysis of 74 kb of DNA located at the right end of the 330-kb
RT   chlorella virus PBCV-1 genome.";
RL   VIROLOGY 237:360-377(1997).
DR   EMBL; U42580; G2447075; -.
SQ   SEQUENCE   313 AA;  37455 MW;  C2AB370D CRC32;

  Query Match             53.2%;  Score 41.5;  DB 11;  Length 313;
  Best Local Similarity   46.7%;  Pred. No. 10;
  Matches    7;  Conservative    4;  Mismatches    3;  Indels    1;  Gaps    1;

Qy       1 RVCAAGRGETCYRTV 15
           :||:  ||: ||| :
Db     161 HVCSC-RGDICYRNI 174


RESULT   4
O18003
ID   O18003      PRELIMINARY;      PRT;   440 AA.
AC   O18003;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   R17.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BARLOW K.;
RL   SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z92809; E1186508; -.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 1.
DR   PFAM; PF00090; tsp_1; 1.
SQ   SEQUENCE   440 AA;  50400 MW;  04D449FB CRC32;


  Query Match             52.6%;  Score 41;  DB 3;  Length 440;
  Best Local Similarity   63.6%;  Pred. No. 16;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RVCAAGRGETC 11
           |:| ||| :||
Db      25 RLCCAGRNNTC 35


RESULT   5
O75331
ID   O75331      PRELIMINARY;      PRT;   303 AA.
AC   O75331;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CATHEPSIN Z PRECURSOR.
GN   CTSZ.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PROSTATE;
RX   MEDLINE; 98307916.
RA   SANTAMARIA I., VELASCO G., PENDAS A.M., FUEYO A., LOPEZ-OTIN C.;
RT   "Cathepsin Z, a novel human cysteine proteinase with a short
RT   propeptide domain and a unique chromosomal location.";
RL   J. BIOL. CHEM. 273:16816-16823(1998).
DR   EMBL; AF032906; G3294548; -.
KW   SIGNAL.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        62    303       CATHEPSIN Z.
SQ   SEQUENCE   303 AA;  33882 MW;  7485A4BA CRC32;


  Query Match             51.3%;  Score 40;  DB 2;  Length 303;
  Best Local Similarity   66.7%;  Pred. No. 18;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       7 RGETCYRTV 15
           ||:|||| :
Db      29 RGQTCYRPL 37


RESULT   6
P94354
ID   P94354      PRELIMINARY;      PRT;   218 AA.
AC   P94354;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HOMOLOGOUS TO MANY TRANSCRIPTIONAL ACTIVATOR PROTEINS.
GN   YXJL.
OS   BACILLUS SUBTILIS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BGSC 1A1;
RX   MEDLINE; 97124196.
RA   YOSHIDA K., SHINDO K., SANO H., SEKI S., FUJIMURA M., YANAI N.,
RA   MIWA Y., FUJITA Y.;
RT   "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT   containing the lic and cel loci, and creation of a 177 kb contig
```

```
RT   "covering the gnt-sacXY region.";
RL   MICROBIOLOGY 142:3113-3123(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA   AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA   BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA   BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA   CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA   DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA   ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D., FRITZ C.,
RA   FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N., GHIM S.Y.,
RA   GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G., GUISEPPI G.,
RA   GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A., HILBERT H.,
RA   HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L., JORIS B.,
RA   KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C., KOBAYASHI Y.,
RA   KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M., KURITA K., LAPIDUS A.,
RA   LARDINOIS S., LAUBER J., LAZAREVIC V., LEE S.M., LEVINE A., LIU H.,
RA   MASUDA S., MAUEL C., MEDIGUE C., MEDINA N., MELLADO R.P., MIZUNO M.,
RA   MOESTL D., NAKAI S., NOBACK M., NOONE D., O'REILLY M., OGAWA K.,
RA   OGIWARA A., OUDEGA B., PARK S.H., PARRO V., POHL T.M., PORTETELLE D.,
RA   PORWOLLIK S., PRESCOTT A.M., PRESECAN E., PUJIC P., PURNELLE B.,
RA   RAPOPORT G., REY M., REYNOLDS S., RIEGER M., RIVOLTA C., ROCHA E.,
RA   ROCHE B., ROSE M., SADAIE Y., SATO T., SCANLAN E., SCHLEICH S.,
RA   SCHROETER R., SCOFFONE F., SEKIGUCHI J., SEKOWSKA A., SEROR S.J.,
RA   SERROR P., SHIN B.S., SOLDO B., SOROKIN A., TACCONI E., TAKAGI T.,
RA   TAKAHASHI H., TAKEMARU K., TAKEUCHI M., TAMAKOSHI A., TANAKA T.,
RA   TERPSTRA P., TOGNONI A., TOSATO V., UCHIYAMA S., VANDENBOL M.,
RA   VANNIER F., VASSAROTTI A., VIARI A., WAMBUTT R., WEDLER E., WEDLER H.,
RA   WEITZENEGGER T., WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K.,
RA   YASUMOTO K., YATA K., YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E.,
RT   YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   NATURE 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; D83026; D1012380; -.
DR   EMBL; Z99123; E1186390; -.
DR   PROSITE; PS00622; HTH_LUXR_FAMILY; 1.
DR   PFAM; PF00072; response_reg; 1.
DR   PFAM; PF00196; GerE; 1.
SQ   SEQUENCE   218 AA;  24245 MW;  1D8B7BD8 CRC32;
```

```
  Query Match             50.0%;  Score 39;  DB 9;  Length 218;
  Best Local Similarity   66.7%;  Pred. No. 20;
  Matches    8;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 AAGRGETCYRTV 15
          ||:|||: :|||
Db    121 AAARGEAIFRTV 132
```

```
RESULT   7
Q88593
ID   Q88593      PRELIMINARY;      PRT;   139 AA.
AC   Q88593;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PROTEASE (FRAGMENT).
OS   THEILER MURINE ENCEPHALOMYELITIS VIRUS.
OC   VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC   CARDIOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87061197.
RA   OZDEN S., TANGY F., CHAMORRO M., BRAHIC M.;
RT   "Theiler's virus genome is closely related to that of
RT   encephalomyocarditis virus, the prototype cardiovirus.";
RL   J. VIROL. 60:1163-1165(1986).
DR   EMBL; M14703; E12025; -.
KW   PROTEASE.
FT   NON_TER       1      1
SQ   SEQUENCE   139 AA;  15070 MW;  7309B868 CRC32;
```

```
  Query Match             48.7%;  Score 38;  DB 11;  Length 139;
  Best Local Similarity   55.6%;  Pred. No. 20;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 AGRGETCYR 13
          ::| :||||
Db     38 PARNDTCYR 46
```

```
RESULT   8
O38024
ID   O38024      PRELIMINARY;      PRT;    92 AA.
AC   O38024;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 10.3 KD PROTEIN.
OS   POTATO VIRUS M (PVM).
OC   VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CARLAVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IDAHO;
RA   CAVILEER T.D., CORSINI D.L., BERGER P.H.;
RL   SUBMITTED (SEP-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF023877; G2547199; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   92 AA;  10319 MW;  9C955295 CRC32;
```

```
  Query Match             47.4%;  Score 37;  DB 11;  Length 92;
  Best Local Similarity   61.5%;  Pred. No. 21;
  Matches    8;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 RVCAAGRGETCYR 13
          |  ||||   |||
Db     50 RAIAAGRCHRCYR 62
```

```
RESULT   9
P93694
ID   P93694      PRELIMINARY;      PRT;   488 AA.
AC   P93694;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PHEROPHORIN III (FRAGMENT).
OS   VOLVOX CARTERI.
OC   EUKARYOTA; VIRIDIPLANTAE; CHLOROPHYTA; CHLOROPHYCEAE; VOLVOCALES;
OC   VOLVOCACEAE; VOLVOX.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK10 (FEMALE);
RX   MEDLINE; 96073432.
RA   GODL K., HALLMANN A., RAPPEL A., SUMPER M.;
RT   "Pherophorins: a family of extracellular matrix glycoproteins from
RT   Volvox structurally related to the sex-inducing pheromone.";
RL   PLANTA 196:781-787(1995).
DR   EMBL; X82446; E305342; -.
FT   NON_TER       1      1
SQ   SEQUENCE   488 AA;  51465 MW;  71A5542F CRC32;
```

```
  Query Match             50.0%;  Score 39;  DB 8;  Length 488;
```

Best Local Similarity 50.0%; Pred. No. 39;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Qy      2 VCAAGRGETCYRTV 15
          || :| || || ::
Db    458 VCKGGIGEQCYVSI 471
```

RESULT 10
O44335
ID O44335 PRELIMINARY; PRT; 208 AA.
AC O44335;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE CHEMOATTRACTIVE GLYCOPROTEIN ES20.
GN ES20.
OS LUMBRICUS TERRESTRIS (COMMON EARTHWORM).
OC EUKARYOTA; METAZOA; ANNELIDA; CLITELLATA; OLIGOCHAETA; HAPLOTAXIDA;
OC LUMBRICINA; LUMBRICIDAE; LUMBRICUS.
RN [1]
RP SEQUENCE FROM N.A.
RA WANG D., LIU W., HALPERN M., CHEN P.;
RL J. BIOL. CHEM. 0:0-0(1997).
DR EMBL; AF019234; G2738885; -.
SQ SEQUENCE 208 AA; 22954 MW; 5C4365F6 CRC32;

Query Match 46.8%; Score 36.5; DB 3; Length 208;
Best Local Similarity 36.8%; Pred. No. 50;
Matches 7; Conservative 5; Mismatches 2; Indels 5; Gaps 1;

```
Qy      2 VCAAG-----RGETCYRTV 15
          ||::|      ||:||: :
Db     17 VCPPGFTYLPAGESCYKVI 35
```

RESULT 11
Q94248
ID Q94248 PRELIMINARY; PRT; 1127 AA.
AC Q94248;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CODED FOR BY C. ELEGANS CDNA YK6G11.3.
GN H03E18.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA BENTLEY D.;
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; U67947; G1519650; -.
SQ SEQUENCE 1127 AA; 124650 MW; C6C909EA CRC32;

Query Match 51.3%; Score 40; DB 3; Length 1127;
Best Local Similarity 60.0%; Pred. No. 52;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy      6 GRGETCYRTV 15
          |:|  ||||:
Db    217 GKGPVCYRTI 226
```

RESULT 12
P89137
ID P89137 PRELIMINARY; PRT; 179 AA.
AC P89137;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ORF 5.
OS SIMIAN HEMORRHAGIC FEVER VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LVR 42-0/M6941;
RA ZENG L., GODENY E.K., METHVEN S.L., BRINTON M.A.;
RL VIROLOGY 207:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LVR 42-0/M6941;
RX MEDLINE; 97361760.
RA SMITH S.L., WANG X., GODENY E.K.;
RT "Sequence of the 3' end of the simian hemorrhagic fever virus
RT genome.";
RL GENE 191:205-210(1997).
DR EMBL; U63121; G1762563; -.
SQ SEQUENCE 179 AA; 19655 MW; ED8A5791 CRC32;

Query Match 46.2%; Score 36; DB 11; Length 179;
Best Local Similarity 30.8%; Pred. No. 53;
Matches 4; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```
Qy      3 CAAGRGETCYRTV 15
          |:| :   |:| :
Db     25 CSANQTHICFRNI 37
```

RESULT 13
O14522
ID O14522 PRELIMINARY; PRT; 472 AA.
AC O14522;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0283 (EC 3.1.3.48) (FRAGMENT).
GN KIAA0283.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA OHARA O., NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N.,
RA NOMURA N.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AB006621; D1023828; -.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PFAM; PF00102; Y_phosphatase; 2.

KW HYDROLASE.
FT NON_TER 1 1
SQ SEQUENCE 472 AA; 54092 MW; 039D0105 CRC32;

Query Match 48.7%; Score 38; DB 2; Length 472;
Best Local Similarity 40.0%; Pred. No. 55;
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```
Qy      1 RVCAAGRGETCYRTV 15
          |:|  :| :  || |
Db    349 RICNMARPQDGYRIV 363
```

RESULT 14
Q20876
ID Q20876 PRELIMINARY; PRT; 137 AA.
AC Q20876;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COSMID F56D3.
GN F56D3.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WILCOX L.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; U28942; G861331; -.
SQ SEQUENCE 137 AA; 15158 MW; A239ADFB CRC32;

Query Match 44.9%; Score 35; DB 3; Length 137;
Best Local Similarity 33.3%; Pred. No. 62;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 RVCAAGRGETCYRTV 15
          |||:  |: :|: ::
Db     55 RVCSIRRSFACFSSI 69
```

RESULT 15
O52666
ID O52666 PRELIMINARY; PRT; 1420 AA.
AC O52666;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE CORE PROTEIN.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC45;
RA WANG Y.-D., ZHAO S., HILL C.W.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AF044501; G2920634; -.
SQ SEQUENCE 1420 AA; 159215 MW; 76C21D24 CRC32;

Query Match 51.3%; Score 40; DB 9; Length 1420;
Best Local Similarity 77.8%; Pred. No. 63;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy      5 AGRGETCYR 13
          ||| |:|||
Db    364 AGRPESCYR 372
```

Search completed: May 3, 1999, 17:54:25
Job time: 645 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 16:05:43 ; Search time 18.06 Seconds
(without alignments)
22.293 Million cell updates/sec

Title: US-09-037-460-2_COPY_55_69
Perfect score: 78
Sequence: 1 RVCAAGRGETCYRTV 15

Scoring table: PAM150

Searched: 74019 seqs, 26840295 residues

Database : SwissProt_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	51.3	115	1	H2AL_HUMAN	P98176 homo sapien
2	42	53.8	607	1	ALBU_HORSE	P35747 equus cabal
3	39	50.0	218	1	YXJL_BACSU	P55184 bacillus su
4	36	46.2	83	1	NPAB_LOCMI	P10776 locusta mig
5	33.5	42.9	60	1	CX6_NAJAT	P80245 naja atra (
6	31	39.7	25	1	CXOB_CONMA	P05485 conus magus
7	33.5	42.9	81	1	CX8_NAJAT	P49123 naja atra (
8	33.5	42.9	81	1	CXN_NAJAT	P49124 naja atra (
9	38	48.7	728	1	HGF_HUMAN	P14210 homo sapien
10	30	38.5	19	1	HMD_METWO	P32441 methanobate
11	29.5	37.8	17	1	TAC1_TACGI	P23684 tachypleus
12	29.5	37.8	17	1	TAC3_TACGI	P18252 tachypleus
13	36	46.2	351	1	NOV_CHICK	P28686 gallus gall
14	36	46.2	353	1	NOV_COTJA	P42642 coturnix co
15	36	46.2	354	1	NOV_MOUSE	Q64299 mus musculu
16	36	46.2	357	1	NOV_HUMAN	P48745 homo sapien
17	35	44.9	244	1	RS6_BRAFL	O01727 branchiosto
18	33.5	42.9	138	1	PA2A_VIPAA	P00626 vipera ammo
19	33.5	42.9	138	1	PA2B_VIPAA	P14424 vipera ammo
20	33.5	42.9	138	1	PA2C_VIPAA	P11407 vipera ammo
21	37	47.4	721	1	YHC6_CAEEL	Q23316 caenorhabdi
22	37	47.4	728	1	HGF_MOUSE	Q08048 mus musculu
23	37	47.4	728	1	HGF_RAT	P17945 rattus norv
24	31.5	40.4	60	1	CX1_NAJNA	P01447 naja naja (
25	31.5	40.4	60	1	CX2_NAJKA	P01445 naja naja k
26	31.5	40.4	60	1	CX2_NAJNA	P01440 naja naja (
27	31.5	40.4	60	1	CX2_NAJOX	P01441 naja oxiana
28	31.5	40.4	60	1	CX3_NAJKA	P01446 naja naja k
29	31.5	40.4	60	1	CX3_NAJNA	P24780 naja naja (
30	31.5	40.4	60	1	CX5T_NAJAT	P07525 naja atra (
31	31.5	40.4	60	1	CX5_NAJKA	P24779 naja naja k
32	36	46.2	472	1	VTDB_MOUSE	P21614 mus musculu
33	36	46.2	559	1	UROT_MOUSE	P11214 mus musculu
34	34	43.6	232	1	YFJ0_YEAST	P43600 saccharomyc
35	36	46.2	608	1	ALBU_RAT	P02770 rattus norv
36	32	41.0	98	1	GAS1_ARATH	P46689 arabidopsis
37	31.5	40.4	81	1	CX1B_NAJAT	Q98956 naja atra (
38	31.5	40.4	81	1	CX1C_NAJAT	Q98957 naja atra (
39	31.5	40.4	81	1	CX1D_NAJAT	Q98958 naja atra (
40	31.5	40.4	81	1	CX1_NAJAT	P01449 naja atra (
41	31.5	40.4	81	1	CX1_NAJSP	Q02454 naja sputat
42	31.5	40.4	81	1	CX2_NAJAT	P01442 naja atra (
43	31.5	40.4	81	1	CX3A_NAJAT	Q98959 naja atra (
44	31.5	40.4	81	1	CX3B_NAJAT	Q98960 naja atra (
45	31.5	40.4	81	1	CX3D_NAJAT	Q98961 naja atra (

ALIGNMENTS

RESULT 1
H2AL_HUMAN
ID H2AL_HUMAN STANDARD; PRT; 115 AA.
AC P98176;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 12.7 KD HISTONE H2A RELATED PROTEIN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96133299.
RA NAYLOR J.A., BUCK D., GREEN P.M., WILLIAMSON H., BENTLEY D.,
RA GIANNELLI F.;
RL HUM. MOL. GENET. 4:1217-1224(1995).
CC -!- SIMILARITY: TO HISTONE H2A.
DR EMBL; X86012; G976083; -.
DR PROSITE; PS00046; HISTONE_H2A; FALSE_NEG.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 115 AA; 12697 MW; 729A09E6 CRC32;

Query Match 51.3%; Score 40; DB 1; Length 115;
Best Local Similarity 75.0%; Pred. No. 2.7;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       4 AAGRGETCYRTV 15
           |:||| || |||
Db      13 AGGRGRTCSRTV 24
```

RESULT 2
ALBU_HORSE
ID ALBU_HORSE STANDARD; PRT; 607 AA.
AC P35747;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SERUM ALBUMIN PRECURSOR.
GN ALB.
OS EQUUS CABALLUS (HORSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PERISSODACTYLA.
RN [1]
RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RC TISSUE=LIVER;
RX MEDLINE; 93345495.
RA HO J.X., HOLOWACHUK E.W., NORTON E.J., TWIGG P.D., CARTER D.C.;
RL EUR. J. BIOCHEM. 215:205-212(1993).
CC -!- FUNCTION: SERUM ALBUMIN, THE MAIN PROTEIN OF PLASMA, HAS A GOOD
CC BINDING CAPACITY FOR WATER, CA++, NA+, K+, FATTY ACIDS, HORMONES,
CC BILIRUBIN AND DRUGS. ITS MAIN FUNCTION IS THE REGULATION OF THE
CC COLLOIDAL OSMOTIC PRESSURE OF BLOOD.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: PLASMA.
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
DR EMBL; X74045; G399672; -.
DR PIR; S34053; ABHOS.
DR PROSITE; PS00212; ALBUMIN; 3.
KW PLASMA; METAL-BINDING; LIPID-BINDING; ALBUMIN; REPEAT; SIGNAL;
KW COPPER.
FT SIGNAL 1 18 BY SIMILARITY.
FT PROPEP 19 24 BY SIMILARITY.

```
FT   CHAIN        25    607       SERUM ALBUMIN.
FT   REPEAT       28    202       1.
FT   REPEAT      221    393       2.
FT   REPEAT      412    591       3.
FT   METAL        27     27       COPPER (BY SIMILARITY).
FT   DISULFID     77     86
FT   DISULFID     99    115
FT   DISULFID    114    125
FT   DISULFID    147    192
FT   DISULFID    191    200
FT   DISULFID    223    269
FT   DISULFID    268    276
FT   DISULFID    288    302
FT   DISULFID    301    312
FT   DISULFID    339    384
FT   DISULFID    383    392
FT   DISULFID    415    461
FT   DISULFID    460    471
FT   DISULFID    484    500
FT   DISULFID    499    510
FT   DISULFID    537    582
FT   DISULFID    581    590
SQ   SEQUENCE   607 AA;  68598 MW;  040EBFB8 CRC32;


  Query Match             53.8%;  Score 42;  DB 1;  Length 607;
  Best Local Similarity   46.7%;  Pred. No. 5;
  Matches    7;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 RVCAAGRGETCYRTV 15
           : || :   :|||||
Db     382 KCCAEADPPACYRTV 396


RESULT   3
YXJL_BACSU
ID   YXJL_BACSU     STANDARD;      PRT;   218 AA.
AC   P55184;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN GALE-PEPT INTERGENIC REGION.
GN   YXJL.
OS   BACILLUS SUBTILIS.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   SCHROEGEL O., KRISPIN O., ALLMANSBERGER R.;
RL   FEMS MICROBIOL. LETT. 145:341-348(1996).
CC   -!- FUNCTION: PROBABLE MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM
CC       YXJM/YXJL.
CC   -!- PTM: PHOSPHORYLATED BY YXJM (PROBABLE).
CC   -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC       REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
CC   -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
DR   EMBL; X99339; E253894; -.
DR   SUBTILIST; BG11895; YXJL.
DR   PROSITE; PS00622; HTH_LUXR_FAMILY; 1.
KW   HYPOTHETICAL PROTEIN; SENSORY TRANSDUCTION; PHOSPHORYLATION;
KW   TRANSCRIPTION REGULATION; DNA-BINDING.
FT   DOMAIN       1      ?       RECEIVER DOMAIN.
FT   MOD_RES     58     58       PHOSPHORYLATION (BY SIMILARITY).
FT   DNA_BIND    174    193      H-T-H MOTIF (POTENTIAL).
SQ   SEQUENCE   218 AA;  24259 MW;  91CAF541 CRC32;


  Query Match             50.0%;  Score 39;  DB 1;  Length 218;
  Best Local Similarity   66.7%;  Pred. No. 6.7;
  Matches    8;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 AAGRGETCYRTV 15
           ||:|||: :|||
Db     121 AAARGEAIFRTV 132


RESULT   4
NPAB_LOCMI
ID   NPAB_LOCMI     STANDARD;      PRT;    83 AA.
AC   P10776;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE   NEUROPARSIN A (NPA) (CONTAINS: NEUROPARSIN B (NPB)).
OS   LOCUSTA MIGRATORIA (MIGRATORY LOCUST).
OC   EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; ORTHOPTERA.
RN   [1]
RP   SEQUENCE.
RA   GIRARDIE J., HUET J.C., PERNOLLET J.C.;
RL   INSECT BIOCHEM. 20:659-666(1990).
RN   [2]
RP   SEQUENCE OF 6-83.
RC   TISSUE=NEUROPORA CARDIACA;
RX   MEDLINE; 89171328.
RA   GIRARDIE J., GIRARDIE A., HUET J.-C., PERNOLLET J.-C.;
RL   FEBS LETT. 245:4-8(1989).
CC   -!- FUNCTION: NEUROSPARINS ARE MULTIFUNCTIONAL NEUROHORMONES: THEY
CC       INHIBIT THE EFFECTS OF JUVENILE HORMONE, STIMULATE FLUID
CC       REABSORPTION OF ISOLATED RECTA AND INDUCES AN INCREASE IN
CC       HEMOLYMPH LIPID AND TREHALOSE LEVELS.
CC   -!- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
DR   PIR; A60361; A60361.
KW   HORMONE; NEUROPEPTIDE.
FT   PEPTIDE       1     83       NEUROPARSIN A.
FT   PEPTIDE       6     83       NEUROPARSIN B.
SQ   SEQUENCE   83 AA;  8765 MW;  8E9BE0B8 CRC32;


  Query Match             46.2%;  Score 36;  DB 1;  Length 83;
  Best Local Similarity   54.5%;  Pred. No. 9.5;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 RVCAAGRGETC 11
           :||| | |: |
Db      31 KVCAKGPGDKC 41


RESULT   5
CX6_NAJAT
ID   CX6_NAJAT      STANDARD;      PRT;    60 AA.
AC   P80245;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   CYTOTOXIN 6 (CARDIOTOXIN 6) (CTX6).
OS   NAJA ATRA (CHINESE COBRA).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; REPTILIA;
OC   LEPIDOSAURIA; SERPENTES.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=VENOM;
RX   MEDLINE; 94251167.
RA   HUNG C.C., WU S.-H., CHIOU S.-H.;
RL   BIOCHEM. MOL. BIOL. INT. 31:1031-1040(1993).
DR   HSSP; P01444; 2CRS.
DR   PROSITE; PS00272; SNAKE_TOXIN; 1.
KW   VENOM; CYTOTOXIN; CARDIOTOXIN; MULTIGENE FAMILY.
FT   DISULFID      3     21       BY SIMILARITY.
FT   DISULFID     14     38       BY SIMILARITY.
FT   DISULFID     42     53       BY SIMILARITY.
FT   DISULFID     54     59       BY SIMILARITY.
SQ   SEQUENCE   60 AA;  6689 MW;  5F6A8F9C CRC32;


  Query Match             42.9%;  Score 33.5;  DB 1;  Length 60;
```

```
  Best Local Similarity   46.2%;  Pred. No. 19;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 RVCAAGRGETCYR 13
           : |||| : ||:
Db      12 KTCAAGK-NLCYK 23


RESULT   6
CXOB_CONMA
ID   CXOB_CONMA      STANDARD;      PRT;     25 AA.
AC   P05485;
DT   01-NOV-1988 (REL. 09, CREATED)
DT   01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT   01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)
DE   OMEGA-CONOTOXIN MVIIB.
OS   CONUS MAGUS (MAGUS CONE).
OC   EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PROSOBRANCHIA;
OC   NEOGASTROPODA; CONIDAE.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 87299637.
RA   OLIVERA B.M., CRUZ L.J., DE SANTOS V., LECHEMINANT G.W., GRIFFIN D.,
RA   ZEIKUS R.D., MCINTOSH J.M., GALYEAN R., VARGA J., GRAY W.R.,
RL   BIOCHEMISTRY 26:2086-2090(1987).
RN   [2]
RP   REVIEW.
RX   MEDLINE; 89024586.
RA   GRAY W.R., OLIVERA B.M., CRUZ L.J.;
RL   ANNU. REV. BIOCHEM. 57:665-700(1988).
CC   -!- FUNCTION: OMEGA-CONOTOXINS ACT AT PRESYNAPTIC MEMBRANES, THEY BIND
CC       AND BLOCK THE CALCIUM CHANNELS.
DR   PIR; B34115; B34115.
DR   PIR; JH0701; JH0701.
KW   PRESYNAPTIC NEUROTOXIN; CALCIUM CHANNEL INHIBITOR; VENOM;
KW   AMIDATION.
FT   DISULFID      1     16
FT   DISULFID      8     20
FT   DISULFID     15     25
FT   MOD_RES      25     25       AMIDATION.
SQ   SEQUENCE   25 AA;  2626 MW;  D9978077 CRC32;


  Query Match             39.7%;  Score 31;  DB 1;  Length 25;
  Best Local Similarity   55.6%;  Pred. No. 24;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 GRGETCYRT 14
           |:| :| ||
Db       3 GKGASCHRT 11


RESULT   7
CX8_NAJAT
ID   CX8_NAJAT      STANDARD;      PRT;     81 AA.
AC   P49123;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CYTOTOXIN 8 PRECURSOR (CARDIOTOXIN 8) (CTX8).
OS   NAJA ATRA (CHINESE COBRA).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; REPTILIA;
OC   LEPIDOSAURIA; SERPENTES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=VENOM GLAND;
RA   CHANG L.-S., LIN J., WU P.F.;
RL   SUBMITTED (SEP-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z54229; G1000507; -.
DR   PROSITE; PS00272; SNAKE_TOXIN; 1.
KW   VENOM; CYTOTOXIN; CARDIOTOXIN; MULTIGENE FAMILY; SIGNAL.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22     81       CYTOTOXIN 8.
FT   DISULFID     24     42       BY SIMILARITY.
FT   DISULFID     35     59       BY SIMILARITY.
FT   DISULFID     63     74       BY SIMILARITY.
FT   DISULFID     75     80       BY SIMILARITY.
SQ   SEQUENCE   81 AA;  8923 MW;  689435F4 CRC32;


  Query Match             42.9%;  Score 33.5;  DB 1;  Length 81;
  Best Local Similarity   46.2%;  Pred. No. 24;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 RVCAAGRGETCYR 13
           : |||| : ||:
Db      33 KACAAGK-NLCYK 44


RESULT   8
CXN_NAJAT
ID   CXN_NAJAT      STANDARD;      PRT;     81 AA.
AC   P49124;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CYTOTOXIN N PRECURSOR (CARDIOTOXIN N) (CTXN).
OS   NAJA ATRA (CHINESE COBRA).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; REPTILIA;
OC   LEPIDOSAURIA; SERPENTES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=VENOM GLAND;
RX   MEDLINE; 96190679.
RA   CHANG L.-S., WU P.F., LIN J.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 219:116-121(1996).
DR   EMBL; Z54230; G1000509; -.
DR   PROSITE; PS00272; SNAKE_TOXIN; 1.
KW   VENOM; CYTOTOXIN; CARDIOTOXIN; MULTIGENE FAMILY; SIGNAL.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22     81       CYTOTOXIN N.
FT   DISULFID     24     42       BY SIMILARITY.
FT   DISULFID     35     59       BY SIMILARITY.
FT   DISULFID     63     74       BY SIMILARITY.
FT   DISULFID     75     80       BY SIMILARITY.
SQ   SEQUENCE   81 AA;  8952 MW;  08D10592 CRC32;


  Query Match             42.9%;  Score 33.5;  DB 1;  Length 81;
  Best Local Similarity   46.2%;  Pred. No. 24;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 RVCAAGRGETCYR 13
           : |||| : ||:
Db      33 KTCAAGK-NLCYK 44


RESULT   9
HGF_HUMAN
ID   HGF_HUMAN      STANDARD;      PRT;    728 AA.
AC   P14210;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HEPATOCYTE GROWTH FACTOR PRECURSOR (SCATTER FACTOR) (SF)
DE   (HEPATOPOEITIN A).
GN   HGF OR HPTA.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91340155.
RA   SEKI T., HAGIYA M., SHIMONISHI M., NAKAMURA T., SHIMIZU S.;
RL   GENE 102:213-219(1991).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 89392017.
RA   MIYAZAWA K., TSUBOUCHI H., NAKA D., TAKAHASHI K., OKIGAKI M.,
RA   ARAKAKI N., NAKAYAMA H., HIRONO S., SAKIYAMA O., TAKAHASHI K.,
RA   GOHDA E., DAIKUHARA Y., KITAMURA N.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 163:967-973(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEUKOCYTE;
RX   MEDLINE; 91025062.
RA   SEKI T., IHARA I., SUGIMURA A., SHIMONISHI M., NISHIZAWA T.,
RA   ASAMI O., HAGIYA M., NAKAMURA T., SHIMIZU S.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 172:321-327(1990).
RN   [4]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 55-73 AND 495-520.
RC   TISSUE=LIVER;
RX   MEDLINE; 90066676.
RA   NAKAMURA T., NISHIZAWA T., HAGIYA M., SEKI T., SHIMONISHI M.,
RA   SUGIMURA A., TASHIRO K., SHIMIZU S.;
RL   NATURE 342:440-443(1989).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYONIC FIBROBLAST;
RX   MEDLINE; 91334393.
RA   WEIDNER K.M., ARAKAKI N., HARTMANN G., VANDEKERCKHOVE J., WEINGART S.,
RA   RIEDER H., FONATSCH C., TSUBOUCHI H., HISHIDA T., DAIKUHARA Y.,
RA   BIRCHMEIER W.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 88:7001-7005(1991).
RN   [6]
RP   SIGNAL SEQUENCE CLEAVAGE SITE.
RX   MEDLINE; 91207365.
RA   YOSHIYAMA Y., ARAKAKAI N., NAKA D., TAKAHASHI K., HIRONO S., KONDO J.,
RA   NAKAYAMA H., GOHDA E., KITAMURA N., TSUBOUCHI H., ISHII T.,
RA   HISHIDA T., DAIKUHARA Y.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 175:660-667(1991).
RN   [7]
RP   MUTAGENESIS.
RX   MEDLINE; 92331602.
RA   LOKKER N.A., MARK M.R., LUIS E.A., BENNETT G.L., ROBBINS K.A.,
RA   BAKER J.B., GODOWSKI P.J.;
RL   EMBO J. 11:2503-2510(1992).
RN   [8]
RP   STRUCTURE BY NMR OF 31-127.
RX   MEDLINE; 98154323.
RA   ZHOU H., MAZZULLA M.J., KAUFMAN J.D., STAHL S.J., WINGFIELD P.T.,
RA   RUBIN J.S., BOTTARO D.P., BYRD R.A.;
RL   STRUCTURE 6:109-116(1998).
CC   -!- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL
CC       HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPHIC FACTOR, AND ACTS
CC       AS GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES.
CC       IT HAS NO DETECTABLE PROTEASE ACTIVITY.
CC   -!- SUBUNIT: DIMER OF AN ALPHA CHAIN AND A BETA CHAIN LINKED BY A
CC       DISULFIDE BOND.
CC   -!- SIMILARITY: CONTAINS 4 KRINGLE REGIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
DR   EMBL; D90334; G219700; -.
DR   EMBL; D90318; G219700; JOINED.
DR   EMBL; D90319; G219700; JOINED.
DR   EMBL; D90320; G219700; JOINED.
DR   EMBL; D90322; G219700; JOINED.
DR   EMBL; D90323; G219700; JOINED.
DR   EMBL; D90324; G219700; JOINED.
DR   EMBL; D90325; G219700; JOINED.
DR   EMBL; D90326; G219700; JOINED.
DR   EMBL; D90327; G219700; JOINED.
DR   EMBL; D90328; G219700; JOINED.
DR   EMBL; D90329; G219700; JOINED.
DR   EMBL; D90330; G219700; JOINED.
DR   EMBL; D90331; G219700; JOINED.
DR   EMBL; D90332; G219700; JOINED.
DR   EMBL; D90333; G219700; JOINED.
DR   EMBL; M29145; G306846; -.
DR   EMBL; M60718; G184032; -.
DR   EMBL; X16323; G32082; -.
DR   EMBL; M73239; G337936; -.
DR   EMBL; M73240; G337938; -.
DR   PIR; JH0579; JH0579.
DR   PIR; S06794; S06794.
DR   PDB; 2HGF; 24-JUN-98.
DR   MIM; 142409; -.
DR   PROSITE; PS00021; KRINGLE_1; 4.
DR   PROSITE; PS50070; KRINGLE_2; 4.
KW   GROWTH FACTOR; KRINGLE; GLYCOPROTEIN; SERINE PROTEASE HOMOLOG;
KW   SIGNAL; 3D-STRUCTURE.
FT   SIGNAL       1     31
FT   CHAIN       32    494       HEPATOCYTE GROWTH FACTOR ALPHA CHAIN.
FT   CHAIN      495    728       HEPATOCYTE GROWTH FACTOR BETA CHAIN.
FT   MOD_RES     32     32       PYRROLIDONE CARBOXYLIC ACID.
FT   DOMAIN      32    127       PAP.
FT   DOMAIN     128    206       KRINGLE 1.
FT   DOMAIN     211    288       KRINGLE 2.
FT   DOMAIN     305    383       KRINGLE 3.
FT   DOMAIN     391    469       KRINGLE 4.
FT   DOMAIN     495    728       SERINE PROTEASE-LIKE.
FT   DISULFID    70     96
FT   DISULFID    74     84
FT   DISULFID   487    604       INTERCHAIN (BY SIMILARITY).
FT   CARBOHYD   294    294       POTENTIAL.
FT   CARBOHYD   402    402       POTENTIAL.
FT   CARBOHYD   566    566       POTENTIAL.
FT   CARBOHYD   653    653       POTENTIAL.
FT   CONFLICT    32     33       QR -> HK (IN REF. 4).
FT   CONFLICT    78     78       K -> N (IN REF. 4).
FT   CONFLICT   293    293       M -> V (IN REF. 4).
FT   CONFLICT   300    300       L -> M (IN REF. 4).
FT   CONFLICT   317    317       V -> A (IN REF. 4).
FT   CONFLICT   336    336       E -> K (IN REF. 4).
FT   CONFLICT   387    387       H -> N (IN REF. 4).
FT   CONFLICT   416    416       D -> N (IN REF. 4).
FT   CONFLICT   505    505       I -> V (IN REF. 4).
FT   CONFLICT   509    509       V -> I (IN REF. 4).
FT   CONFLICT   558    558       D -> E (IN REF. 4).
FT   CONFLICT   561    561       C -> R (IN REF. 4).
FT   CONFLICT   595    595       S -> N (IN REF. 4).
SQ   SEQUENCE   728 AA;  83133 MW;  5182013A CRC32;


  Query Match          48.7%;  Score 38;  DB 1;  Length 728;
  Best Local Similarity   45.5%;  Pred. No. 28;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CAAGRGETCYR 13
           |  ::|: |||
Db     383 CDMSHGQDCYR 393


RESULT  10
HMD_METWO
ID   HMD_METWO      STANDARD;      PRT;    19 AA.
AC   P32441;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   H(2)-FORMING N5,N10-METHYLENETETRAHYDROMETHANOPTERIN DEHYDROGENASE
DE   (EC 1.12.99.-) (H2-DEPENDENT METHYLENE-H4MPT DEHYDROGENASE)
DE   (FRAGMENT).
GN   HMD.
OS   METHANOBATERIUM WOLFEI.
OC   ARCHAEBACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;
OC   METHANOBACTERIACEAE.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92394151.
```

```
RA   ZIRNGIBL C., VAN DONGEN W., SCHWOERER B., VON BUENAU R.,
RA   RICHTER M., KLEIN A., THAUER R.K.;
RL   EUR. J. BIOCHEM. 208:511-520(1992).
CC   -!- CATALYTIC ACTIVITY: 5,10-METHYLENETETRAHYDROMETHANOPTERIN +
CC       H(+) = 5,10-METHENYLTETRAHYDROMETHANOPTERIN + H(2).
CC   -!- COFACTOR: ZINC (POSSIBLE).
CC   -!- PATHWAY: INVOLVED IN METHANOGENESIS.
CC   -!- SUBUNIT: HOMODIMER.
KW   OXIDOREDUCTASE; METHANOGENESIS; ZINC.
FT   NON_TER       1      1
SQ   SEQUENCE   19 AA;  1911 MW;  749D8DDC CRC32;


  Query Match           38.5%;  Score 30;  DB 1;  Length 19;
  Best Local Similarity  71.4%;  Pred. No. 28;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       8 GETCYRT 14
           |  ||||
Db       7 GAGCYRT 13


RESULT  11
TAC1_TACGI
ID   TAC1_TACGI     STANDARD;      PRT;    17 AA.
AC   P23684;
DT   01-NOV-1991 (REL. 20, CREATED)
DT   01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TACHYPLESIN I.
OS   TACHYPLEUS GIGAS (SOUTHEAST ASIAN HORSESHOE CRAB), AND CARCINOSCORPIUS
OS   ROTUNDICAUDA (SOUTHEAST ASIAN HORSESHOE CRAB).
OC   EUKARYOTA; METAZOA; ARTHROPODA; CHELICERATA; XIPHOSURA.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=T.GIGAS, AND C.ROTUNDICAUDA;
RX   MEDLINE; 91035357.
RA   MUTA T., FUJIMOTO T., NAKAJIMA H., IWANAGA S.;
RL   J. BIOCHEM. 108:261-266(1990).
CC   -!- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.
DR   PIR; A38824; A38824.
DR   PIR; JX0124; JX0124.
KW   ANTIBIOTIC; AMIDATION.
FT   DISULFID      3     16
FT   DISULFID      7     12
FT   MOD_RES      17     17       AMIDATION.
SQ   SEQUENCE   17 AA;  2269 MW;  7B171B7B CRC32;


  Query Match           37.8%;  Score 29.5;  DB 1;  Length 17;
  Best Local Similarity  53.8%;  Pred. No. 31;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    1;

Qy       1 RVCAAGRGETCYR 13
           |||  |    |||
Db       5 RVCYRG---ICYR 14


RESULT  12
TAC3_TACGI
ID   TAC3_TACGI     STANDARD;      PRT;    17 AA.
AC   P18252;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TACHYPLESIN III.
OS   TACHYPLEUS GIGAS (SOUTHEAST ASIAN HORSESHOE CRAB).
OC   EUKARYOTA; METAZOA; ARTHROPODA; CHELICERATA; XIPHOSURA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 91035357.
RA   MUTA T., FUJIMOTO T., NAKAJIMA H., IWANAGA S.;
RL   J. BIOCHEM. 108:261-266(1990).
CC   -!- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND
CC       GRAM-POSITIVE BACTERIA.
CC   -!- TISSUE SPECIFICITY: HEMOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.
DR   PIR; JX0125; JX0125.
KW   ANTIBIOTIC; AMIDATION.
FT   DISULFID      3     16       BY SIMILARITY.
FT   DISULFID      7     12       BY SIMILARITY.
FT   MOD_RES      17     17       AMIDATION.
SQ   SEQUENCE   17 AA;  2241 MW;  68E08384 CRC32;


  Query Match           37.8%;  Score 29.5;  DB 1;  Length 17;
  Best Local Similarity  53.8%;  Pred. No. 31;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    1;

Qy       1 RVCAAGRGETCYR 13
           |||  |    |||
Db       5 RVCYRG---ICYR 14


RESULT  13
NOV_CHICK
ID   NOV_CHICK      STANDARD;      PRT;   351 AA.
AC   P28686;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN PRECURSOR.
GN   NOV.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BROWN LEGHORN;
RX   MEDLINE; 92107157.
RA   MALOISEL V., MARTINERIE C., DAMBRINE G., PLASSIART G., BRISAC M.,
RA   CROCHET J., PERBAL B.;
RL   MOL. CELL. BIOL. 12:10-21(1992).
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION. ITS OVEREXPRESSION IS ASSOCIATED WITH
CC       TUMORIGENESIS AND EXPRESSION OF A N-TERMINAL-TRUNCATED VERSION
CC       OF NOV GENE IN CHICKEN EMBRYONIC FIBROBLASTS (CEF) IS SUFFICIENT
CC       TO INDUCE THE TRANSFORMATION OF CEF IN VITRO.
CC   -!- TISSUE SPECIFICITY: BRAIN AND HEART, AND AT A LOWER LEVEL IN
CC       MUSCLE AND INTESTINE, IN THE EMBRYO. LUNG AND LESS SO IN BRAIN AND
CC       SPLEEN, IN ADULT CHICKEN.
CC   -!- DEVELOPMENTAL STAGE: MAV1-INDUCED NEPHROBLASTOMAS EXPRESS A HIGH
CC       LEVEL OF NOV GENE WHOSE TRANSCRIPTION IS NORMALLY ARRESTED IN
CC       ADULT KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; X59284; G63703; -.
DR   PIR; S20078; S20078.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    351       NOV PROTEIN.
FT   DOMAIN      104    170       VWFC.
FT   DOMAIN      258    332       CTCK.
FT   DISULFID    258    295       BY SIMILARITY.
FT   DISULFID    275    309       BY SIMILARITY.
FT   DISULFID    286    325       BY SIMILARITY.
FT   DISULFID    289    327       BY SIMILARITY.
FT   DISULFID    294    331       BY SIMILARITY.
FT   CARBOHYD    274    274       POTENTIAL.
SQ   SEQUENCE   351 AA;  38268 MW;  C7044065 CRC32;
```

```
  Query Match             46.2%;  Score 36;  DB 1;  Length 351;
  Best Local Similarity   70.0%;  Pred. No. 32;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 VCAAGRGETC 11
           |||  |||:|
Db      62 VCARQRGESC 71


RESULT  14
NOV_COTJA
ID   NOV_COTJA      STANDARD;      PRT;   353 AA.
AC   P42642;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN PRECURSOR.
GN   NOV.
OS   COTURNIX COTURNIX JAPONICA (JAPANESE QUAIL).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WEISKIRCHEN R., BISTER K.;
RL   SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; U13063; G532697; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    353       NOV PROTEIN.
FT   DOMAIN      106    172       VWFC.
FT   DOMAIN      260    334       CTCK.
FT   DISULFID    260    297       BY SIMILARITY.
FT   DISULFID    277    311       BY SIMILARITY.
FT   DISULFID    288    327       BY SIMILARITY.
FT   DISULFID    291    329       BY SIMILARITY.
FT   DISULFID    296    333       BY SIMILARITY.
FT   CARBOHYD    276    276       POTENTIAL.
SQ   SEQUENCE   353 AA;  38667 MW;  C4F5928D CRC32;


  Query Match             46.2%;  Score 36;  DB 1;  Length 353;
  Best Local Similarity   70.0%;  Pred. No. 32;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 VCAAGRGETC 11
           |||  |||:|
Db      64 VCARQRGESC 73


RESULT  15
NOV_MOUSE
ID   NOV_MOUSE      STANDARD;      PRT;   354 AA.
AC   Q64299;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN HOMOLOG PRECURSOR (NOVH).
GN   NOV.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV, AND ICR; TISSUE=BRAIN;
RA   SNAITH M.R., NATARAJAN D., TAYLOR L.B., CHOI C.P., MARTINERIE C.,
RA   PERBAL B., SCHOFIELD P.N., BOULTER C.A.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE; 96204003.
RA   MARTINERIE C., CHEVALIER G., RAUSCHER F.J. III, PERBAL B.;
RL   ONCOGENE 12:1479-1492(1996).
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; X97863; E1181580; -.
DR   EMBL; Y09257; E281106; -.
DR   EMBL; X96585; E228599; -.
DR   MGD; MGI:109185; NOV.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    354       NOV PROTEIN HOMOLOG.
FT   DOMAIN      102    168       VWFC.
FT   DOMAIN      261    335       CTCK.
FT   DISULFID    261    298       BY SIMILARITY.
FT   DISULFID    278    312       BY SIMILARITY.
FT   DISULFID    289    328       BY SIMILARITY.
FT   DISULFID    292    330       BY SIMILARITY.
FT   DISULFID    297    334       BY SIMILARITY.
FT   CARBOHYD     91     91       POTENTIAL.
FT   CARBOHYD    277    277       POTENTIAL.
SQ   SEQUENCE   354 AA;  38928 MW;  65A4FF0E CRC32;


  Query Match             46.2%;  Score 36;  DB 1;  Length 354;
  Best Local Similarity   70.0%;  Pred. No. 32;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 VCAAGRGETC 11
           |||  |||:|
Db      60 VCARQRGESC 69


Search completed: May  3, 1999, 16:05:43
Job time: 4002 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 16:05:42 ; Search time 18.06 Seconds
(without alignments)
22.293 Million cell updates/sec

Title: US-09-037-460-2_COPY_30_44
Perfect score: 74
Sequence: 1 QHCDSSECKSSPRCK 15

Scoring table: PAM150

Searched: 74019 seqs, 26840295 residues

Database : SwissProt_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	59.5	64	1	MTCU_HELPO	P55947 helix pomat
2	40	54.1	58	1	MT2_CALSI	P55950 callinectes
3	38	51.4	74	1	MT_CRAVI	P23038 crassostrea
4	42	56.8	488	1	FA10_HUMAN	P00742 homo sapien
5	37	50.0	61	1	MTC_YEAST	P07215 saccharomyc
6	36	48.6	57	1	MT2_SCYSE	P02806 scylla serr
7	44	59.5	2524	1	NOTC_XENLA	P21783 xenopus lae
8	44	59.5	2531	1	NTC1_MOUSE	Q01705 mus musculu
9	36	48.6	66	1	MT3_MALDO	O24059 malus domes
10	37	50.0	111	1	YLM1_CAEEL	P34375 caenorhabdi
11	35	47.3	58	1	MT_POTPO	P55952 potamon pot
12	39	52.7	400	1	B3AR_MOUSE	P25962 mus musculu
13	39	52.7	400	1	B3AR_RAT	P26255 rattus norv
14	39	52.7	410	1	TRA8_PSESH	P24607 pseudomonas
15	35	47.3	80	1	MT1_COFAR	P43396 coffea arab
16	38	51.4	335	1	PIE1_CAEEL	Q94131 caenorhabdi
17	42	56.8	2444	1	NTC1_HUMAN	P46531 homo sapien
18	38	51.4	379	1	ADH1_HORVU	P05336 hordeum vul
19	38	51.4	379	1	ADH1_PENAM	P14219 pennisetum
20	36	48.6	155	1	CONO_LYMST	Q00945 lymnaea sta
21	38	51.4	414	1	THIK_YARLI	Q05493 yarrowia li
22	34	45.9	64	1	MT3_PRUAV	O48951 prunus aviu
23	39.5	53.4	837	1	MUCL_RAT	P98089 rattus norv
24	34	45.9	65	1	MT_PARLI	P80367 paracentrot
25	34	45.9	68	1	NXL1_ASPSC	P25670 aspidelaps
26	38	51.4	466	1	FA7_HUMAN	P08709 homo sapien
27	34	45.9	74	1	MT1_CAEEL	P17511 caenorhabdi
28	35	47.3	125	1	NEU1_MOUSE	P35454 mus musculu
29	35	47.3	125	1	NEU1_RAT	P01179 rattus norv
30	37	50.0	337	1	FSA_SHEEP	P31514 ovis aries
31	37	50.0	338	1	FSA_HUMAN	P19883 homo sapien
32	37	50.0	344	1	FSA_BOVIN	P50291 bos taurus
33	37	50.0	344	1	FSA_MOUSE	P47931 mus musculu
34	37	50.0	344	1	FSA_PIG	P10669 sus scrofa
35	37	50.0	344	1	FSA_RAT	P21674 rattus norv
36	40	54.1	1429	1	LI12_CAEEL	P14585 caenorhabdi
37	34.5	46.6	110	1	IBP_TURRS	P00993 red sea tur
38	38	51.4	603	1	FA12_CAVPO	Q04962 cavia porce
39	36	48.6	243	1	TLPH_ARATH	P50699 arabidopsis
40	36	48.6	243	1	TPX1_HUMAN	P16562 homo sapien
41	41	55.4	2531	1	NTC1_RAT	Q07008 rattus norv
42	33	44.6	63	1	MT3_ACTCH	P43389 actinidia c
43	33	44.6	64	1	MTA_STRPU	P04734 strongyloce
44	37	50.0	416	1	FA9_BOVIN	P00741 bos taurus
45	33	44.6	67	1	MTA_SPHGR	Q26497 sphaerechin

ALIGNMENTS

RESULT 1
MTCU_HELPO
ID MTCU_HELPO STANDARD; PRT; 64 AA.
AC P55947;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COPPER-METALLOTHIONEIN (CU-MT).
OS HELIX POMATIA (ROMAN SNAIL) (EDIBLE SNAIL).
OC EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PULMONATA; STYLOMMATOPHORA;
OC HELICIDAE.
RN [1]
RP SEQUENCE.
RC TISSUE=MANTLE;
RX MEDLINE; 97373947.
RA DALLINGER R., BERGER B., HUNZIKER P.E., KAEGI J.H.R.;
RL NATURE 388:237-238(1997).
CC -!- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR
CC SEQUESTRATION OF TOXIC METAL IONS AND REGULATION OF ESSENTIAL
CC TRACE ELEMENTS. THIS ISOFORM BINDS EXCLUSIVELY COPPER.
CC -!- DOMAIN: 14 CYSTEINE RESIDUES ARE ARRANGED IN C-X-C GROUPS. THESE
CC ARE THOUGHT TO BE THE METAL-BINDING SITES IN OTHER
CC METALLOTHIONEINS.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 IN METALLOTHIONEIN SUPERFAMILY.
KW METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION; COPPER;
KW ACETYLATION.
FT MOD_RES 1 1 ACETYLATION.
FT METAL 7 7
FT METAL 11 11
FT METAL 16 16
FT METAL 18 18
FT METAL 22 22
FT METAL 24 24
FT METAL 28 28
FT METAL 30 30
FT METAL 33 33
FT METAL 36 36
FT METAL 38 38
FT METAL 43 43
FT METAL 45 45
FT METAL 49 49
FT METAL 55 55
FT METAL 57 57
FT METAL 61 61
FT METAL 63 63
SQ SEQUENCE 64 AA; 6205 MW; 1E294035 CRC32;

Query Match 59.5%; Score 44; DB 1; Length 64;
Best Local Similarity 38.5%; Pred. No. 1.2;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```
Qy      2 HCDSSECKSSPRC 14
          || : :|| :::|
Db     37 HCSNDDCKCGSQC 49
```

RESULT 2
MT2_CALSI
ID MT2_CALSI STANDARD; PRT; 58 AA.
AC P55950;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE METALLOTHIONEIN-II (MT-II) (MT-IIB/MT-IIA).
OS CALLINECTES SAPIDUS (BLUE CRAB).

```
OC   EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; MALACOSTRACA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 96033062.
RA   BROUWER M., ENGHILD J., HOEXUM-BROUWER T., THOGERSEN I., TRUNCALI A.;
RL   BIOCHEM. J. 311:617-622(1995).
CC   -!- FUNCTION: BINDS SIX DIVALENT METAL IONS. KNOWN TO BIND COPPER AND
CC       CADMIUM.
CC   -!- SIMILARITY: BELONGS TO FAMILY 3 IN METALLOTHIONEIN SUPERFAMILY.
KW   METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION; COPPER; CADMIUM.
FT   DOMAIN        1     29       BETA.
FT   DOMAIN       30     58       ALPHA.
FT   METAL         5      5       CLUSTER B.
FT   METAL         6      6       CLUSTER B.
FT   METAL        10     10       CLUSTER B.
FT   METAL        12     12       CLUSTER B.
FT   METAL        17     17       CLUSTER B.
FT   METAL        21     21       CLUSTER B.
FT   METAL        23     23       CLUSTER B.
FT   METAL        26     26       CLUSTER B.
FT   METAL        28     28       CLUSTER B.
FT   METAL        31     31       CLUSTER A.
FT   METAL        34     34       CLUSTER A.
FT   METAL        38     38       CLUSTER A.
FT   METAL        40     40       CLUSTER A.
FT   METAL        46     46       CLUSTER A.
FT   METAL        50     50       CLUSTER A.
FT   METAL        54     54       CLUSTER A.
FT   METAL        56     56       CLUSTER A.
FT   METAL        57     57       CLUSTER A.
FT   VARIANT       1      1       MISSING (IN MT-IIA).
SQ   SEQUENCE   58 AA;  6287 MW;  097C8AA0 CRC32;


  Query Match             54.1%;  Score 40;  DB 1;  Length 58;
  Best Local Similarity   46.2%;  Pred. No. 4.4;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRCK 15
           |  :|||:: :||
Db      12 CKEGECKTGCKCK 24


RESULT   3
MT_CRAVI
ID   MT_CRAVI       STANDARD;      PRT;    74 AA.
AC   P23038;
DT   01-NOV-1991 (REL. 20, CREATED)
DT   01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   METALLOTHIONEIN (MT).
OS   CRASSOSTREA VIRGINICA (EASTERN OYSTER).
OC   EUKARYOTA; METAZOA; MOLLUSCA; PELECYPODA (BIVALVIA); PTERIOMORPHIA;
OC   OSTREOIDA; OSTREIDAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91363394.
RA   UNGER M.E., CHEN T.T., MURPHY C.M., VESTLING M.M., FENSELAU C.,
RA   ROESIJADI G.;
RL   BIOCHIM. BIOPHYS. ACTA 1074:371-377(1991).
RN   [2]
RP   SEQUENCE OF 1-27.
RC   TISSUE=GILL;
RX   MEDLINE; 89372898.
RA   ROESIJADI G., KIELLAND S., KLERKS P.;
RL   ARCH. BIOCHEM. BIOPHYS. 273:403-413(1989).
RN   [3]
RP   SEQUENCE OF 1-24, AND ACETYLATION.
RC   TISSUE=GILL;
RX   MEDLINE; 91291855.
RA   ROESIJADI G., VESTLING M.M., MURPHY C.M., KLERKS P.L., FENSELAU C.C.;
RL   BIOCHIM. BIOPHYS. ACTA 1074:230-236(1991).
CC   -!- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR
CC       SEQUESTRATION OF TOXIC METAL IONS.
CC   -!- INDUCTION: BY CADMIUM.
CC   -!- PTM: THIS MT OCCURS IN BOTH ACETYLATED AND NON-ACETYLATED FORMS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 IN METALLOTHIONEIN SUPERFAMILY.
DR   EMBL; X59862; G288278; -.
DR   PIR; S05645; S05645.
DR   PIR; S17156; S17156.
DR   HSSP; P02805; 1DME.
KW   METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION; CADMIUM;
KW   ACETYLATION.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION (PARTIAL).
SQ   SEQUENCE   74 AA;  7213 MW;  587F7336 CRC32;


  Query Match             51.4%;  Score 38;  DB 1;  Length 74;
  Best Local Similarity   46.2%;  Pred. No. 10;
  Matches    6;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRCK 15
           | :: || :| ||
Db      18 CPATGCKCGPGCK 30


RESULT   4
FA10_HUMAN
ID   FA10_HUMAN     STANDARD;      PRT;   488 AA.
AC   P00742;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   COAGULATION FACTOR X PRECURSOR (EC 3.4.21.6) (STUART FACTOR).
GN   F10.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91216473.
RA   MESSIER T.L., PITTMAN D.D., LONG G.L., KAUFMAN R.J., CHURCH W.R.;
RL   GENE 99:291-294(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87026600.
RA   LEYTUS S.P., FOSTER D.C., KURACHI K., DAVIE E.W.;
RL   BIOCHEMISTRY 25:5098-5102(1986).
RN   [3]
RP   SEQUENCE OF 13-488 FROM N.A.
RX   MEDLINE; 85216545.
RA   FUNG M.R., HAY C.W., MCGILLIVRAY R.T.A.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 82:3591-3595(1985).
RN   [4]
RP   SEQUENCE OF 19-488 FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 86221713.
RA   KAUL R.K., HILDEBRAND B., ROBERTS S., JAGADEESWARAN P.;
RL   GENE 41:311-314(1986).
RN   [5]
RP   SEQUENCE OF 41-179.
RX   MEDLINE; 83257207.
RA   MCMULLEN B.A., FUJIKAWA K., KISIEL W., SASAGAWA T., HOWALD W.N.,
RA   KWA E.Y., WEINSTEIN B.;
RL   BIOCHEMISTRY 22:2875-2884(1983).
RN   [6]
RP   SEQUENCE OF 115-488 FROM N.A.
RX   MEDLINE; 84222026.
RA   LEYTUS S.P., CHUNG D.W., KISIEL W., KURACHI K., DAVIE E.W.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 81:3699-3702(1984).
RN   [7]
RP   SEQUENCE OF 183-234, AND CARBOHYDRATE-BINDING SITES.
RX   MEDLINE; 94062825.
RA   INOUE K., MORITA T.;
RL   EUR. J. BIOCHEM. 218:153-163(1993).
```

```
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 86-179 AND 235-278.
RX   MEDLINE; 93360277.
RA   PADMANABHAN K., PADMANABHAN K.P., TULINSKY A., PARK C.H., BODE W.,
RA   HUBER R., BLANKENSHIP D.T., CARDIN A.D., KISIEL W.;
RL   J. MOL. BIOL. 232:947-966(1993).
CC   -!- FUNCTION: FACTOR XA IS A VITAMIN K-DEPENDENT GLYCOPROTEIN THAT
CC       CONVERTS PROTHROMBIN TO THROMBIN IN THE PRESENCE OF FACTOR VA,
CC       CA++, AND PHOSPHOLIPID DURING BLOOD CLOTTING.
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-|-THR AND THEN
CC       ARG-|-ILE BONDS IN PROTHROMBIN TO FORM THROMBIN.
CC   -!- SUBUNIT: THE TWO CHAINS ARE FORMED FROM A SINGLE-CHAIN PRECURSOR
CC       BY THE EXCISION OF TWO ARG RESIDUES AND ARE HELD TOGETHER BY 1 OR
CC       MORE DISULFIDE BONDS.
CC   -!- PTM: THE VITAMIN K-DEPENDENT, ENZYMATIC CARBOXYLATION OF SOME
CC       GLUTAMIC ACID RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND
CC       CALCIUM.
CC   -!- PTM: N- AND O-GLYCOSYLATED.
CC   -!- PTM: THE ACTIVATION PEPTIDE IS CLEAVED BY FACTOR IXA (IN THE
CC       INTRINSIC PATHWAY), OR BY FACTOR VIIA (IN THE EXTRINSIC PATHWAY).
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
DR   EMBL; K03194; G182841; -.
DR   EMBL; M57285; G182390; -.
DR   EMBL; L29433; G182831; -.
DR   EMBL; L00390; G182831; JOINED.
DR   EMBL; L00391; G182831; JOINED.
DR   EMBL; L00392; G182831; JOINED.
DR   EMBL; L00393; G182831; JOINED.
DR   EMBL; L00394; G182831; JOINED.
DR   EMBL; L00395; G182831; JOINED.
DR   EMBL; L00396; G182831; JOINED.
DR   EMBL; M22613; G180336; -.
DR   PIR; A00924; EXHU.
DR   PIR; A25853; A25853.
DR   PIR; A24478; A24478.
DR   PDB; 1HCG; 08-MAY-95.
DR   PDB; 1FAX; 29-OCT-97.
DR   MIM; 134530; -.
DR   MIM; 227600; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   GLYCOPROTEIN; HYDROLASE; SERINE PROTEASE; PLASMA; BLOOD COAGULATION;
KW   GAMMA-CARBOXYGLUTAMIC ACID; HYDROXYLATION; CALCIUM-BINDING; VITAMIN K;
KW   SIGNAL; ZYMOGEN; EGF-LIKE DOMAIN; REPEAT; 3D-STRUCTURE.
FT   SIGNAL       1      ?
FT   PROPEP       ?     40
FT   CHAIN       41    179       FACTOR X LIGHT CHAIN.
FT   CHAIN      183    488       FACTOR X HEAVY CHAIN.
FT   PROPEP     183    234       ACTIVATION PEPTIDE.
FT   CHAIN      235    488       ACTIVATED FACTOR XA, HEAVY CHAIN.
FT   DOMAIN      86    122       EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     125    165       EGF-LIKE 2.
FT   DOMAIN     235    488       CATALYTIC.
FT   MOD_RES     46     46       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     47     47       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     54     54       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     56     56       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     59     59       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     60     60       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     65     65       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     66     66       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     69     69       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     72     72       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     79     79       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES    103    103       HYDROXYLATION.
FT   CARBOHYD   199    199
FT   CARBOHYD   211    211
FT   CARBOHYD   221    221
FT   CARBOHYD   231    231
FT   ACT_SITE   276    276       CHARGE RELAY SYSTEM.
FT   ACT_SITE   322    322       CHARGE RELAY SYSTEM.
FT   ACT_SITE   419    419       CHARGE RELAY SYSTEM.
FT   DISULFID    90    101
FT   DISULFID    95    110
FT   DISULFID   112    121
FT   DISULFID   129    140
FT   DISULFID   136    149
FT   DISULFID   151    164
FT   DISULFID   172    342       INTERCHAIN.
FT   DISULFID   241    246
FT   DISULFID   261    277
FT   DISULFID   390    404
FT   DISULFID   415    443
FT   CONFLICT   285    288       KVRV -> E (IN REF. 4 AND 6).
FT   CONFLICT   442    442       G -> S (IN REF. 3).
SQ   SEQUENCE   488 AA;  54731 MW;  49BD04F3 CRC32;

  Query Match           56.8%;  Score 42;  DB 1;  Length 488;
  Best Local Similarity  33.3%;  Pred. No. 10;
  Matches    5;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 QHCDSSECKSSPRCK 15
           ::|::| | :  :||
Db      88 DQCETSPCQNQGKCK 102


RESULT   5
MTC_YEAST
ID   MTC_YEAST      STANDARD;      PRT;    61 AA.
AC   P07215;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   METALLOTHIONEIN (CU-MT) PRECURSOR (COPPER CHELATIN).
GN   CUP1 OR MTH1 OR YHR053C OR YHR055C.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84119482.
RA   KARIN M., NAJARIAN R.C., HASLINGER A., VALENZUELA P., WELCH J.,
RA   FOGEL S.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 81:337-341(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84221953.
RA   BUTT T.R., STERNBERG E.J., GORMAN J.A., CLARK P., HAMER D.,
RA   ROSENBERG M., CROOKE S.T.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 81:3332-3336(1984).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91203809.
RA   JEYAPRAKASH A., WELCH J.W., FOGEL S.;
RL   MOL. GEN. GENET. 225:363-368(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 94378003.
RA   JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA   DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
RA   KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
RA   LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
RA   NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
RA   VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
RA   VAUDIN M.;
RL   SCIENCE 265:2077-2082(1994).
RN   [5]
RP   SEQUENCE.
```

```
RX   MEDLINE; 86033949.
RA   WINGE D.R., NIELSON K.B., GRAY W.R., HAMER D.H.;
RL   J. BIOL. CHEM. 260:14464-14470(1985).
RN   [6]
RP   X-RAY ABSORPTION STUDIES.
RX   MEDLINE; 88228044.
RA   GEORGE G.N., BYRD J., WINGE D.R.;
RL   J. BIOL. CHEM. 263:8199-8203(1988).
RN   [7]
RP   STRUCTURE BY NMR OF 9-48.
RX   MEDLINE; 96159028.
RA   PETERSON C.W., NARULA S.S., ARMITAGE I.M.;
RL   FEBS LETT. 379:85-93(1996).
CC   -!- FUNCTION: PROTECTS THE CELL AGAINST COPPER TOXICITY BY TIGHTLY
CC       CHELATING COPPER IONS. CAN BINDS 8 COPPER IONS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 12 IN METALLOTHIONEIN SUPERFAMILY.
DR   EMBL; K02204; G171338; -.
DR   EMBL; U00061; G487953; -.
DR   EMBL; U00061; G487951; -.
DR   PIR; A29373; A29373.
DR   PIR; S14049; S14049.
DR   PIR; S46703; S46703.
DR   PIR; S46705; S46705.
DR   PDB; 1AOO; 24-DEC-97.
DR   PDB; 1AQQ; 24-DEC-97.
DR   PDB; 1AQR; 24-DEC-97.
DR   PDB; 1AQS; 24-DEC-97.
DR   SGD; L0000436; CUP1.
KW   METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION; COPPER;
KW   3D-STRUCTURE.
FT   PROPEP          1      8
FT   CHAIN           9     61       METALLOTHIONEIN.
SQ   SEQUENCE   61 AA;  6650 MW;  A7B05F14 CRC32;


  Query Match             50.0%;  Score 37;  DB 1;  Length 61;
  Best Local Similarity   30.8%;  Pred. No. 12;
  Matches    4;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 HCDSSECKSSPRC 14
           :|: : ||:: :|
Db      16 QCQCGSCKNNEQC 28


RESULT   6
MT2_SCYSE
ID   MT2_SCYSE      STANDARD;      PRT;    57 AA.
AC   P02806;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   METALLOTHIONEIN-II (MT-II).
OS   SCYLLA SERRATA (MUD CRAB).
OC   EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; MALACOSTRACA.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 82142340.
RA   LERCH K., AMMER D., OLAFSON R.W.;
RL   J. BIOL. CHEM. 257:2420-2426(1982).
CC   -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC       RESIDUES THAT BIND VARIOUS HEAVY METALS. CLASS I MTS IN MARINE
CC       CRUSTACEA ARE INVOLVED IN THE SEQUESTRATION OF ELEVATED LEVELS
CC       OF HEAVY-METAL IONS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 3 IN METALLOTHIONEIN SUPERFAMILY.
DR   PIR; A03284; SMKD2S.
DR   HSSP; P02805; 1DMC.
KW   METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION.
FT   DOMAIN          1     28       BETA.
FT   DOMAIN         29     57       ALPHA.
FT   METAL           4      4       CLUSTER B.
FT   METAL           5      5       CLUSTER B.
FT   METAL           9      9       CLUSTER B.
FT   METAL          11     11       CLUSTER B.
FT   METAL          16     16       CLUSTER B.
FT   METAL          20     20       CLUSTER B.
FT   METAL          22     22       CLUSTER B.
FT   METAL          25     25       CLUSTER B.
FT   METAL          27     27       CLUSTER B.
FT   METAL          30     30       CLUSTER A.
FT   METAL          33     33       CLUSTER A.
FT   METAL          37     37       CLUSTER A.
FT   METAL          39     39       CLUSTER A.
FT   METAL          45     45       CLUSTER A.
FT   METAL          49     49       CLUSTER A.
FT   METAL          53     53       CLUSTER A.
FT   METAL          55     55       CLUSTER A.
FT   METAL          56     56       CLUSTER A.
SQ   SEQUENCE   57 AA;  6109 MW;  4DEEE236 CRC32;


  Query Match             48.6%;  Score 36;  DB 1;  Length 57;
  Best Local Similarity   41.7%;  Pred. No. 16;
  Matches    5;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRC 14
           |  :|||:: :|
Db      11 CKEGECKTGCKC 22


RESULT   7
NOTC_XENLA
ID   NOTC_XENLA     STANDARD;      PRT;  2524 AA.
AC   P21783;
DT   01-MAY-1991 (REL. 18, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN   XOTCH.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90385285.
RA   COFFMAN C., HARRIS W., KINTNER C.;
RL   SCIENCE 249:1438-1441(1990).
RN   [2]
RP   REVISIONS TO 1759-1782.
RA   KINTNER C.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
DR   EMBL; M33874; G1364263; -.
DR   PIR; A35844; A35844.
DR   HSSP; P20693; 1ESL.
DR   PROSITE; PS00010; ASX_HYDROXYL; 23.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 29.
DR   PROSITE; PS01187; EGF_CA; 21.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL          1     19       POTENTIAL.
FT   CHAIN          20   2524       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT   DOMAIN         20   1728       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     1729   1750       POTENTIAL.
FT   DOMAIN       1751   2524       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN         20     57       EGF-LIKE 1.
FT   DOMAIN         58     99       EGF-LIKE 2.
FT   DOMAIN        102    140       EGF-LIKE 3.
FT   DOMAIN        141    177       EGF-LIKE 4.
FT   DOMAIN        179    215       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN        217    254       EGF-LIKE 6.
FT   DOMAIN        256    292       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
```

FT	DOMAIN	294	332	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	334	370	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	371	409	EGF-LIKE 10.
FT	DOMAIN	411	449	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	451	487	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	489	525	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	527	563	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	565	600	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	602	638	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	640	675	EGF-LIKE 17.
FT	DOMAIN	677	713	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	715	750	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	752	788	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	790	826	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	828	866	EGF-LIKE 22.
FT	DOMAIN	868	904	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	906	942	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	944	980	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	982	1018	EGF-LIKE 26.
FT	DOMAIN	1020	1056	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1058	1094	EGF-LIKE 28.
FT	DOMAIN	1096	1142	EGF-LIKE 29.
FT	DOMAIN	1144	1180	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1182	1218	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1220	1264	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1266	1304	EGF-LIKE 33.
FT	DOMAIN	1306	1346	EGF-LIKE 34.
FT	DOMAIN	1347	1383	EGF-LIKE 35.
FT	DOMAIN	1386	1424	EGF-LIKE 36.
FT	DOMAIN	1441	1560	3 X LIN/NOTCH REPEATS.
FT	REPEAT	1441	1478	LIN/NOTCH 1.
FT	REPEAT	1479	1520	LIN/NOTCH 2.
FT	REPEAT	1521	1560	LIN/NOTCH 3.
FT	DOMAIN	1871	2083	6 X ANK MOTIF REPEATS.
FT	DISULFID	22	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	128	BY SIMILARITY.
FT	DISULFID	130	139	BY SIMILARITY.
FT	DISULFID	145	156	BY SIMILARITY.
FT	DISULFID	150	165	BY SIMILARITY.
FT	DISULFID	167	176	BY SIMILARITY.
FT	DISULFID	183	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	504	BY SIMILARITY.
FT	DISULFID	498	513	BY SIMILARITY.
FT	DISULFID	515	524	BY SIMILARITY.
FT	DISULFID	531	542	BY SIMILARITY.
FT	DISULFID	536	551	BY SIMILARITY.
FT	DISULFID	553	562	BY SIMILARITY.
FT	DISULFID	569	579	BY SIMILARITY.
FT	DISULFID	574	588	BY SIMILARITY.
FT	DISULFID	590	599	BY SIMILARITY.
FT	DISULFID	606	617	BY SIMILARITY.
FT	DISULFID	611	626	BY SIMILARITY.
FT	DISULFID	628	637	BY SIMILARITY.
FT	DISULFID	644	654	BY SIMILARITY.
FT	DISULFID	649	663	BY SIMILARITY.
FT	DISULFID	665	674	BY SIMILARITY.
FT	DISULFID	681	692	BY SIMILARITY.
FT	DISULFID	686	701	BY SIMILARITY.
FT	DISULFID	703	712	BY SIMILARITY.
FT	DISULFID	719	729	BY SIMILARITY.
FT	DISULFID	724	738	BY SIMILARITY.
FT	DISULFID	740	749	BY SIMILARITY.
FT	DISULFID	756	767	BY SIMILARITY.
FT	DISULFID	761	776	BY SIMILARITY.
FT	DISULFID	778	787	BY SIMILARITY.
FT	DISULFID	794	805	BY SIMILARITY.
FT	DISULFID	799	814	BY SIMILARITY.
FT	DISULFID	816	825	BY SIMILARITY.
FT	DISULFID	832	843	BY SIMILARITY.
FT	DISULFID	837	854	BY SIMILARITY.
FT	DISULFID	856	865	BY SIMILARITY.
FT	DISULFID	872	883	BY SIMILARITY.
FT	DISULFID	877	892	BY SIMILARITY.
FT	DISULFID	894	903	BY SIMILARITY.
FT	DISULFID	910	921	BY SIMILARITY.
FT	DISULFID	915	930	BY SIMILARITY.
FT	DISULFID	932	941	BY SIMILARITY.
FT	DISULFID	986	997	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1008	1017	BY SIMILARITY.
FT	DISULFID	1024	1035	BY SIMILARITY.
FT	DISULFID	1029	1044	BY SIMILARITY.
FT	DISULFID	1046	1055	BY SIMILARITY.
FT	DISULFID	1062	1073	BY SIMILARITY.
FT	DISULFID	1067	1082	BY SIMILARITY.
FT	DISULFID	1084	1093	BY SIMILARITY.
FT	DISULFID	1100	1121	BY SIMILARITY.
FT	DISULFID	1115	1130	BY SIMILARITY.
FT	DISULFID	1132	1141	BY SIMILARITY.
FT	DISULFID	1148	1159	BY SIMILARITY.
FT	DISULFID	1153	1168	BY SIMILARITY.
FT	DISULFID	1170	1179	BY SIMILARITY.
FT	DISULFID	1186	1197	BY SIMILARITY.
FT	DISULFID	1191	1206	BY SIMILARITY.
FT	DISULFID	1208	1217	BY SIMILARITY.
FT	DISULFID	1224	1243	BY SIMILARITY.
FT	DISULFID	1237	1252	BY SIMILARITY.
FT	DISULFID	1254	1263	BY SIMILARITY.
FT	DISULFID	1270	1283	BY SIMILARITY.
FT	DISULFID	1275	1292	BY SIMILARITY.
FT	DISULFID	1294	1303	BY SIMILARITY.
FT	DISULFID	1310	1321	BY SIMILARITY.
FT	DISULFID	1315	1333	BY SIMILARITY.
FT	DISULFID	1335	1344	BY SIMILARITY.
FT	DISULFID	1351	1362	BY SIMILARITY.
FT	DISULFID	1356	1371	BY SIMILARITY.
FT	DISULFID	1373	1382	BY SIMILARITY.
FT	DISULFID	1390	1401	BY SIMILARITY.
FT	DISULFID	1395	1412	BY SIMILARITY.
FT	DISULFID	1414	1423	BY SIMILARITY.
FT	CARBOHYD	462	462	POTENTIAL.
FT	CARBOHYD	887	887	POTENTIAL.
FT	CARBOHYD	958	958	POTENTIAL.
FT	CARBOHYD	1178	1178	POTENTIAL.
FT	CARBOHYD	1487	1487	POTENTIAL.
FT	CARBOHYD	1508	1508	POTENTIAL.
FT	CARBOHYD	1584	1584	POTENTIAL.

```
SQ   SEQUENCE   2524 AA;  275123 MW;  E63584B1 CRC32;

  Query Match             59.5%;  Score 44;  DB 1;  Length 2524;
  Best Local Similarity   53.8%;  Pred. No. 17;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRCK 15
           | :| ||:: |||
Db     832 CAGSPCKNGGRCK 844


RESULT   8
NTC1_MOUSE
ID   NTC1_MOUSE      STANDARD;      PRT;  2531 AA.
AC   Q01705;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
GN   NOTCH1 OR MOTCH.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93194170.
RA   FRANCO DEL AMO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,
RA   COPELAND N.G., GRIDLEY T.;
RL   GENOMICS 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 1551-2170 FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93048835.
RA   FRANCO DEL AMO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,
RA   GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.;
RL   DEVELOPMENT 115:737-744(1992).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC   -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
DR   EMBL; Z11886; G288503; -.
DR   MGD; MGI:97363; NOTCH1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
KW   DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KW   TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   DOMAIN       19   1725       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1726   1746       POTENTIAL.
FT   DOMAIN     1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       24   1425       36 X EGF-TYPE REPEATS.
FT   DOMAIN     1449   1462       CYS-RICH.
FT   DOMAIN     1445   1562       3 X LIN/NOTCH REPEATS.
FT   REPEAT     1445   1480       LIN/NOTCH 1.
FT   REPEAT     1481   1522       LIN/NOTCH 2.
FT   REPEAT     1523   1562       LIN/NOTCH 3.
FT   DOMAIN     1865   2075       6 X ANK MOTIF REPEATS.
FT   REPEAT     1865   1910       ANK MOTIF 1.
FT   REPEAT     1912   1942       ANK MOTIF 2.
FT   REPEAT     1944   1975       ANK MOTIF 3.
FT   REPEAT     1978   2009       ANK MOTIF 4.
FT   REPEAT     2011   2042       ANK MOTIF 5.
FT   REPEAT     2044   2075       ANK MOTIF 6.
FT   CARBOHYD    888    888       POTENTIAL.
FT   CARBOHYD    959    959       POTENTIAL.
FT   CARBOHYD   1179   1179       POTENTIAL.
FT   CARBOHYD   1241   1241       POTENTIAL.
FT   CARBOHYD   1489   1489       POTENTIAL.
FT   CARBOHYD   1587   1587       POTENTIAL.
SQ   SEQUENCE   2531 AA;  271312 MW;  AD71189B CRC32;

  Query Match             59.5%;  Score 44;  DB 1;  Length 2531;
  Best Local Similarity   58.3%;  Pred. No. 17;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRC 14
           |||: ||:: ||
Db    1063 CDSAPCKNGGRC 1074


RESULT   9
MT3_MALDO
ID   MT3_MALDO       STANDARD;      PRT;    66 AA.
AC   O24059;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   METALLOTHIONEIN-LIKE PROTEIN TYPE 3.
GN   MT2.
OS   MALUS DOMESTICA (APPLE) (MALUS SYLVESTRIS).
OC   EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; ROSALES;
OC   ROSACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   REID S.J., ROSS G.S.;
RL   PHYSIOL. PLANTARUM 100:183-189(1997).
CC   -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC       RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 15 IN METALLOTHIONEIN SUPERFAMILY.
DR   EMBL; U61974; G1655853; -.
KW   METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION.
SQ   SEQUENCE   66 AA;  6930 MW;  C906D6EC CRC32;

  Query Match             48.6%;  Score 36;  DB 1;  Length 66;
  Best Local Similarity   41.7%;  Pred. No. 18;
  Matches    5;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRC 14
           ||: :|  | :|
Db       5 CDNCDCADSTQC 16


RESULT  10
YLM1_CAEEL
ID   YLM1_CAEEL      STANDARD;      PRT;   111 AA.
AC   P34375;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 12.0 KD PROTEIN D2007.1 IN CHROMOSOME III.
GN   D2007.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
```

```
RA   WOHLDMAN P.;
RL   NATURE 368:32-38(1994).
DR   EMBL; L16560; G289668; -.
DR   PIR; S44787; S44787.
DR   WORMPEP; D2007.1; CE00126.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   111 AA;  12037 MW;  4213E9D9 CRC32;


  Query Match             50.0%;  Score 37;  DB 1;  Length 111;
  Best Local Similarity   33.3%;  Pred. No. 19;
  Matches    5;  Conservative    7;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 QHCDSSECKSSPRCK 15
           ::| :: ||:| :|:
Db      91 DQCGNACCKTSEQCR 105


RESULT  11
MT_POTPO
ID   MT_POTPO       STANDARD;      PRT;    58 AA.
AC   P55952;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   METALLOTHIONEIN (MT).
OS   POTAMON POTAMIOS.
OC   EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; NOT YET CLASSIFIED.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=MIDGUT;
RX   MEDLINE; 97079279.
RA   PEDERSEN S.N., PEDERSEN K.L., HOEJRUP P., DEPLEDGE M.H., KNUDSEN J.;
RL   BIOCHEM. J. 319:999-1003(1996).
CC   -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC       RESIDUES THAT BIND VARIOUS HEAVY METALS. CLASS I MTS IN CRUSTACEA
CC       ARE INVOLVED IN THE SEQUESTRATION OF ELEVATED LEVELS OF HEAVY-
CC       METAL IONS.
CC   -!- INDUCTION: BY CADMIUM.
CC   -!- MASS SPECTROMETRY: MW=6156.8; METHOD=MALDI.
CC   -!- SIMILARITY: BELONGS TO FAMILY 3 IN METALLOTHIONEIN SUPERFAMILY.
KW   METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION; CADMIUM.
FT   DOMAIN        1     29       BETA.
FT   DOMAIN       30     58       ALPHA.
FT   METAL         4      4       CLUSTER B.
FT   METAL         5      5       CLUSTER B.
FT   METAL        10     10       CLUSTER B.
FT   METAL        12     12       CLUSTER B.
FT   METAL        17     17       CLUSTER B.
FT   METAL        21     21       CLUSTER B.
FT   METAL        23     23       CLUSTER B.
FT   METAL        26     26       CLUSTER B.
FT   METAL        28     28       CLUSTER B.
FT   METAL        31     31       CLUSTER A.
FT   METAL        34     34       CLUSTER A.
FT   METAL        38     38       CLUSTER A.
FT   METAL        40     40       CLUSTER A.
FT   METAL        46     46       CLUSTER A.
FT   METAL        50     50       CLUSTER A.
FT   METAL        54     54       CLUSTER A.
FT   METAL        56     56       CLUSTER A.
FT   METAL        57     57       CLUSTER A.
SQ   SEQUENCE   58 AA;  6157 MW;  D5F0EED1 CRC32;


  Query Match             47.3%;  Score 35;  DB 1;  Length 58;
  Best Local Similarity   33.3%;  Pred. No. 23;
  Matches    4;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRC 14
           |: : ||:: :|
Db      12 CEEGKCKAGCKC 23


RESULT  12
B3AR_MOUSE
ID   B3AR_MOUSE     STANDARD;      PRT;   400 AA.
AC   P25962;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   BETA-3 ADRENERGIC RECEPTOR.
GN   ADRB3 OR ADRB3R.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS;
RX   MEDLINE; 92037534.
RA   NAHMIAS C., BLIN N., ELALOUF J.M., MATTEI M.-G., STROSBERG A.D.,
RA   EMORINE L.J.;
RL   EMBO J. 10:3721-3727(1991).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RX   MEDLINE; 93279311.
RA   VAN SPRONSEN A., NAHMIAS C., KRIEF S., BRIEND-SUTREN M.-M.,
RA   STROSBERG A.D., EMORINE L.J.;
RL   EUR. J. BIOCHEM. 213:1117-1124(1993).
RN   [3]
RP   SEQUENCE OF 378-400 FROM N.A.
RC   TISSUE=ADIPOSE TISSUE;
RX   MEDLINE; 93125503.
RA   GRANNEMAN J.G., LAHNERS K.N., RAO D.D.;
RL   MOL. PHARMACOL. 42:964-970(1992).
CC   -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC       INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC       PROTEINS. BETA-3 IS INVOLVED IN THE REGULATION OF LIPOLYSIS AND
CC       THERMOGENESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: WHITE AND BROWN ADIPOSE TISSUES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; X72862; G298113; -.
DR   EMBL; X60438; G50110; -.
DR   EMBL; S53290; G263087; -.
DR   PIR; S18548; S18548.
DR   PIR; S32804; S32804.
DR   GCRDB; GCR_0253; -.
DR   GCRDB; GCR_0551; -.
DR   GCRDB; GCR_0708; -.
DR   MGD; MGI:87939; ADRB3.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT   DOMAIN        1     36       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     37     60       1 (POTENTIAL).
FT   DOMAIN       61     69       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     70     88       2 (POTENTIAL).
FT   DOMAIN       89    108       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    109    130       3 (POTENTIAL).
FT   DOMAIN      131    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    175       4 (POTENTIAL).
FT   DOMAIN      176    200       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    201    222       5 (POTENTIAL).
FT   DOMAIN      223    289       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    290    311       6 (POTENTIAL).
FT   DOMAIN      312    323       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    324    344       7 (POTENTIAL).
FT   DOMAIN      345    400       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      8      8       POTENTIAL.
FT   CARBOHYD     26     26       POTENTIAL.
FT   DISULFID    107    186       BY SIMILARITY.
FT   LIPID       358    358       PALMITATE (BY SIMILARITY).
SQ   SEQUENCE   400 AA;  43006 MW;  474A96AC CRC32;
```

```
  Query Match             52.7%;  Score 39;  DB 1;  Length 400;
  Best Local Similarity   54.5%;  Pred. No. 24;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 DSSECKSSPRC 14
           :: || |:|||
Db     182 EAQECHSNPRC 192


RESULT  13
B3AR_RAT
ID   B3AR_RAT       STANDARD;      PRT;   400 AA.
AC   P26255;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   BETA-3 ADRENERGIC RECEPTOR.
GN   ADRB3 OR ADRB3R.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92084710.
RA   MUZZIN P., REVELLI J.-P., KUHNE F., GOCAYNE J.D., MCCOMBIE W.R.,
RA   VENTER J.C., GIACOBINO J.-P., FRASER C.M.;
RL   J. BIOL. CHEM. 266:24053-24058(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92100057.
RA   GRANNEMAN J.G., LAHNERS K.N., CHAUDHRY A.;
RL   MOL. PHARMACOL. 40:895-899(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93178631.
RA   BENSAID M., KAGHAD M., RODRIGUEZ M., LE FUR G., CAPUT D.;
RL   FEBS LETT. 318:223-226(1993).
CC   -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-
CC       INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G
CC       PROTEINS. BETA-3 IS INVOLVED IN THE REGULATION OF LIPOLYSIS AND
CC       THERMOGENESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: WHITE AND BROWN ADIPOSE TISSUES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; M74716; G202766; -.
DR   EMBL; S73473; G241216; -.
DR   EMBL; S56481; G298307; -.
DR   EMBL; S56152; G298309; -.
DR   PIR; A41679; A41679.
DR   PIR; S29808; S29808.
DR   PIR; A53281; A53281.
DR   GCRDB; GCR_0287; -.
DR   GCRDB; GCR_0622; -.
DR   GCRDB; GCR_0732; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT   DOMAIN       1     36       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    37     60       1 (POTENTIAL).
FT   DOMAIN      61     69       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    70     88       2 (POTENTIAL).
FT   DOMAIN      89    108       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   109    130       3 (POTENTIAL).
FT   DOMAIN     131    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   153    175       4 (POTENTIAL).
FT   DOMAIN     176    200       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   201    222       5 (POTENTIAL).
FT   DOMAIN     223    289       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   290    311       6 (POTENTIAL).
FT   DOMAIN     312    323       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   324    344       7 (POTENTIAL).
FT   DOMAIN     345    400       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     8      8       POTENTIAL.
FT   CARBOHYD    26     26       POTENTIAL.
FT   DISULFID   107    186       BY SIMILARITY.
FT   LIPID      358    358       PALMITATE (BY SIMILARITY).
FT   CONFLICT   234    235       LL -> FV (IN REF. 1).
SQ   SEQUENCE   400 AA;  43146 MW;  963B8842 CRC32;


  Query Match             52.7%;  Score 39;  DB 1;  Length 400;
  Best Local Similarity   54.5%;  Pred. No. 24;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 DSSECKSSPRC 14
           :: || |:|||
Db     182 EAQECHSNPRC 192


RESULT  14
TRA8_PSESH
ID   TRA8_PSESH     STANDARD;      PRT;   410 AA.
AC   P24607;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE   TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS801.
OS   PSEUDOMONAS SYRINGAE (PV. PHASEOLICOLA).
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC   PSEUDOMONADACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LR781;
RX   MEDLINE; 91260445.
RA   ROMANTSCHUK M., RICHTER G.Y., MUKHOPADHYAY P., MILLS D.;
RL   MOL. MICROBIOL. 5:617-622(1991).
CC   -!- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION
CC       SEQUENCE.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR VAL-100 IS THE
CC       INITIATOR.
DR   EMBL; X57269; G45831; -.
DR   PIR; S15163; S15163.
KW   TRANSPOSABLE ELEMENT; TRANSPOSITION; DNA-BINDING; DNA RECOMBINATION.
SQ   SEQUENCE   410 AA;  46980 MW;  B02CD432 CRC32;


  Query Match             52.7%;  Score 39;  DB 1;  Length 410;
  Best Local Similarity   33.3%;  Pred. No. 24;
  Matches    4;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRC 14
           |::  |:: | |
Db      78 CNTCHCRACPSC 89


RESULT  15
MT1_COFAR
ID   MT1_COFAR      STANDARD;      PRT;    80 AA.
AC   P43396;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   METALLOTHIONEIN-LIKE PROTEIN 1.
GN   METAL1.
OS   COFFEA ARABICA (COFFEE).
OC   EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC   GENTIANALES; RUBIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. GUATEMALAN; TISSUE=LEAF;
RA   MOISYADI S., STILES J.I.;
RL   SUBMITTED (JUN-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC       RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 15 IN METALLOTHIONEIN SUPERFAMILY.
DR   EMBL; U11423; G508296; -.
```

KW METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION.
SQ SEQUENCE 80 AA; 7914 MW; BB8060AE CRC32;

Query Match 47.3%; Score 35; DB 1; Length 80;
Best Local Similarity 41.7%; Pred. No. 28;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy       4 DSSECKSSPRCK 15
           ::: || :| ||
Db      62 ENGGCKCGPDCK 73
```

Search completed: May 3, 1999, 16:05:43
Job time: 4002 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 17:54:22 ; Search time 47.06 Seconds
(without alignments)
17.585 Million cell updates/sec

Title: US-09-037-460-2_COPY_30_44
Perfect score: 74
Sequence: 1 QHCDSSECKSSPRCK 15

Scoring table: PAM150

Searched: 180763 seqs, 55169189 residues

Database : SPTREMBL_8:*
1: sp_fungi:*
2: sp_human:*
3: sp_invertebrate:*
4: sp_mammal:*
5: sp_mhc:*
6: sp_organelle:*
7: sp_phage:*
8: sp_plant:*
9: sp_bacteria:*
10: sp_rodent:*
11: sp_virus:*
12: sp_vertebrate:*
13: sp_unclassified:*
14: sp_archea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	74	100.0	184	2	Q15330	Q15330 homo sapien
2	59	79.7	184	10	P97682	P97682 rattus norv
3	42	56.8	157	9	Q52156	Q52156 pseudomonas
4	45	60.8	802	12	O57462	O57462 brachydanio
5	38	51.4	63	8	O65885	O65885 rubus idaeu
6	42	56.8	453	8	O48772	O48772 arabidopsis
7	40	54.1	251	3	Q24774	Q24774 enchytraeus
8	39	52.7	167	10	O70432	O70432 meriones un
9	38	51.4	112	9	Q52067	Q52067 enterobacte
10	41	55.4	473	3	Q25464	Q25464 mytilus gal
11	38	51.4	132	2	Q16609	Q16609 homo sapien
12	42	56.8	772	12	Q92070	Q92070 gallus gall
13	43	58.1	1212	12	O42347	O42347 gallus gall
14	39	52.7	218	8	Q40630	Q40630 oryza sativ
15	41	55.4	547	3	O18190	O18190 caenorhabdi
16	36	48.6	63	8	O23938	O23938 fragaria ve
17	39	52.7	238	8	O04364	O04364 oryza sativ
18	36	48.6	67	7	Q38254	Q38254 bacteriopha
19	36	48.6	73	3	O76953	O76953 lumbricus c
20	36	48.6	77	3	Q25163	Q25163 haemadipsa
21	36	48.6	79	3	O76956	O76956 lumbricus r
22	33.5	45.3	27	8	O81530	O81530 mesembryant
23	37.5	50.7	200	9	Q56898	Q56898 yersinia en
24	38	51.4	250	8	Q43612	Q43612 pennisetum
25	37	50.0	168	3	O16895	O16895 caenorhabdi
26	38	51.4	280	8	O24508	O24508 bambusa gla
27	39	52.7	452	2	O60295	O60295 homo sapien
28	35	47.3	79	14	O26799	O26799 methanobact
29	38	51.4	296	8	Q43722	Q43722 pennisetum
30	38	51.4	296	8	Q43805	Q43805 pennisetum
31	38	51.4	296	8	Q43806	Q43806 pennisetum
32	38	51.4	296	8	Q43807	Q43807 pennisetum
33	38	51.4	298	8	Q43808	Q43808 pennisetum
34	38	51.4	301	3	O76917	O76917 drosophila
35	38	51.4	304	1	Q00867	Q00867 fusarium so
36	38	51.4	319	12	Q91376	Q91376 xenopus lae
37	35	47.3	89	3	O45563	O45563 caenorhabdi
38	38	51.4	343	12	Q90844	Q90844 gallus gall
39	38	51.4	378	3	Q24438	Q24438 drosophila
40	38	51.4	379	8	Q41766	Q41766 zea mays (m
41	34.5	46.6	83	10	O70553	O70553 mus musculu
42	34	45.9	71	3	O76957	O76957 lumbricus t
43	38	51.4	434	10	O55139	O55139 mus musculu
44	39	52.7	683	10	O08883	O08883 mus musculu
45	34	45.9	77	3	O76955	O76955 lumbricus r

ALIGNMENTS

RESULT 1
Q15330
ID Q15330 PRELIMINARY; PRT; 184 AA.
AC Q15330;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ESM-1 SECRETORY PROTEIN PRECURSOR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96355375.
RA LASSALLE P.M., MOLET S., JANIN A., VANDER-HEYDEN J.E., TAVERNIER J.,
RA FIERS W., DEVOS R.E., TONNEL A.E.B.;
RT "ESM-1 is a novel human endothelial cell-specific molecule expressed
RT in lung and regulated by cytokines.";
RL J. BIOL. CHEM. 271:20458-20464(1996).
DR EMBL; X89426; E189266; -.
DR PFAM; PF00219; IGFBP; 1.
KW SIGNAL.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 184 ESM-1 SECRETORY PROTEIN.
SQ SEQUENCE 184 AA; 20095 MW; 08D109DF CRC32;

Query Match 100.0%; Score 74; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QHCDSSECKSSPRCK 15
||||||||||||||||
Db 30 QHCDSSECKSSPRCK 44

RESULT 2
P97682
ID P97682 PRELIMINARY; PRT; 184 AA.
AC P97682;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE PG25.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY; TISSUE=PINEAL GLAND;
RA WANG X., BROWNSTEIN M.J., YOUNG W.S.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.

```
DR   EMBL; U80818; G1750198; -.
SQ   SEQUENCE   184 AA;  20075 MW;  3457564C CRC32;


  Query Match             79.7%;  Score 59;  DB 10;  Length 184;
  Best Local Similarity   66.7%;  Pred. No. 0.02;
  Matches   10;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 QHCDSSECKSSPRCK 15
           :|||::||:|| |||
Db      30 EHCDNTECRSSLRCK 44


RESULT   3
Q52156
ID   Q52156      PRELIMINARY;      PRT;   157 AA.
AC   Q52156;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIDEROPHORE REGULATORY GENE PFRA.
GN   PFRA.
OS   PSEUDOMONAS PUTIDA.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC   PSEUDOMONAS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WCS398;
RX   MEDLINE; 95058183.
RA   VENTURI V., OTTEVANGER C., LEONG J., WEISBEEK P.J.;
RT   "Identification and characterization of a siderophore regulatory gene
RT   (pfrA) of Pseudomonas putida WCS358: homology to the alginate
RT   regulatory gene algQ of Pseudomonas aeruginosa.";
RL   MOL. MICROBIOL. 10:63-73(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WCS398;
RA   VENTURI V.;
RL   SUBMITTED (FEB-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; X75771; G455422; -.
SQ   SEQUENCE   157 AA;  18015 MW;  025EDF95 CRC32;


  Query Match             56.8%;  Score 42;  DB 9;  Length 157;
  Best Local Similarity   46.2%;  Pred. No. 7.4;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 QHCDSSECKSSPR 13
           :||: :||| : |
Db     107 DHCEKGECKDPER 119


RESULT   4
O57462
ID   O57462      PRELIMINARY;      PRT;   802 AA.
AC   O57462;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   DELTAA.
GN   DELTAA.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98165392.
RA   APPEL B., EISEN J.S.;
RT   "Regulation of neuronal specification in the zebrafish spinal cord by
RT   Delta function.";
RL   DEVELOPMENT 125:371-380(1998).
DR   EMBL; AF030031; G2809389; -.
DR   PROSITE; PS01186; EGF_2; 8.
KW   GLYCOPROTEIN.
SQ   SEQUENCE   802 AA;  88941 MW;  42F041BD CRC32;


  Query Match             60.8%;  Score 45;  DB 12;  Length 802;
  Best Local Similarity   50.0%;  Pred. No. 9.5;
  Matches    7;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 QHCDSSECKSSPRC 14
           :|| || | :::||
Db     447 DHCSSSPCSNGARC 460


RESULT   5
O65885
ID   O65885      PRELIMINARY;      PRT;    63 AA.
AC   O65885;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   METALLOTHIONEIN-LIKE PROTEIN.
GN   RAS2.
OS   RUBUS IDAEUS.
OC   EUKARYOTA; VIRIDIPLANTAE; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;
OC   TRACHEOPHYTA; EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;
OC   EUDICOTYLEDONS; ROSIDAE; ROSALES; ROSACEAE; RUBUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. GLEN CLOVA; TISSUE=FRUIT;
RA   JONES C.S., DAVIES H.V., MCNICOL R.J., TAYLOR M.A.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ224146; E1294427; -.
SQ   SEQUENCE   63 AA;  6562 MW;  F32B1F6B CRC32;


  Query Match             51.4%;  Score 38;  DB 8;  Length 63;
  Best Local Similarity   41.7%;  Pred. No. 14;
  Matches    5;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRC 14
           ||| :|  :::|
Db       5 CDSCDCSDASQC 16


RESULT   6
O48772
ID   O48772      PRELIMINARY;      PRT;   453 AA.
AC   O48772;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL PROTEIN.
GN   T21L14.13.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;
OC   TRACHEOPHYTA; EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;
OC   EUDICOTYLEDONS; ROSIDAE; CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA   SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AC003033; G2702272; -.
SQ   SEQUENCE   453 AA;  49747 MW;  C339A73A CRC32;


  Query Match             56.8%;  Score 42;  DB 8;  Length 453;
  Best Local Similarity   46.2%;  Pred. No. 17;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy      3 CDSSECKSSPRCK 15
          |::: || :| ||
Db     94 CETGACKYGPTCK 106


RESULT   7
Q24774
ID   Q24774      PRELIMINARY;      PRT;   251 AA.
AC   Q24774;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CYSTEINE-RICH-PROTEIN.
GN   CRP.
OS   ENCHYTRAEUS BUCHHOLZI.
OC   EUKARYOTA; METAZOA; ANNELIDA; CLITELLATA; OLIGOCHAETA; HAPLOTAXIDA;
OC   TUBIFICINA; ENCHYTRAEIDAE; ENCHYTRAEUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95014230.
RA   WILLUHN J., SCHMITT-WREDE H.P., GREVEN H., WUNDERLICH F.;
RT   "cDNA cloning of a cadmium-inducible mRNA encoding a novel
RT   cysteine-rich, non-metallothionein 25-kDa protein in an enchytraeid
RT   earthworm.";
RL   J. BIOL. CHEM. 269:24688-24691(1994).
DR   EMBL; X79344; G488803; -.
SQ   SEQUENCE   251 AA;  25000 MW;  85747E80 CRC32;


  Query Match            54.1%;  Score 40;  DB 3;  Length 251;
  Best Local Similarity  38.5%;  Pred. No. 22;
  Matches    5;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 CDSSECKSSPRCK 15
          ||: :|| :: |:
Db    223 CDNVNCKCGSSCR 235


RESULT   8
O70432
ID   O70432      PRELIMINARY;      PRT;   167 AA.
AC   O70432;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   BETA-3-ADRENERGIC RECEPTOR (FRAGMENT).
OS   MERIONES UNGUICULATUS (MONGOLIAN JIRD).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; GERBILLINAE; MERIONES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FAT;
RA   WANGEMANN P., LIU J., SHIMOZONO M., SCOFIELD M.A.;
RL   SUBMITTED (MAR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF055351; G3025850; -.
FT   NON_TER       1      1
FT   NON_TER     167    167
SQ   SEQUENCE   167 AA;  18324 MW;  B187BF5E CRC32;


  Query Match            52.7%;  Score 39;  DB 10;  Length 167;
  Best Local Similarity  54.5%;  Pred. No. 22;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 DSSECKSSPRC 14
          :: || |:|||
Db      7 EAQECHSNPRC 17


RESULT   9
Q52067
ID   Q52067      PRELIMINARY;      PRT;   112 AA.
AC   Q52067;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   PLASMID PEA3 NITROGEN FIXATION GENES.
GN   NIFI.
OS   ENTEROBACTER AGGLOMERANS.
OG   PLASMID PEA3.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ENTEROBACTER.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KREUTZER R., STEIBL H.D., DAYANANDA S., DIPPE R., HALDA L., BUCK M.,
RA   KLINGMUELLER W.;
RL   (IN) NITROGEN FIXATION, POLSINELLI M., MATERASSI R., VINCENZINI M.,
RL   EDS., PP 24-26, KLUWER ACADEMIC PUBLISHERS, DORTRECHT, BOSTON, LONDON,
RL   (1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   LECHLER A.;
RL   THESIS (1993), UNIVERSITAET BAYREUTH, GERMANY.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   DEUERLING E.;
RL   THESIS (1993), UNIVERSITAET BAYREUTH, GERMANY.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   STEIBL R., STEIBL H.D., SIDDAVATTAM D., KLINGMEULLER W.;
RL   (IN) NEW HORIZONS IN NITROGEN FIXATION, PALACIOS R., MORA J.,
RL   NEWTON W.E., EDS., PP 496-496, KLUWER ACADEMIC PUBLISHERS,
RL   DORTRECHT (1993).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   SCHWICKERATH O.;
RL   THESIS (1993), DIPLOMARBEIT, UNIVERSITAET BAYREUTH, GERMANY.
DR   EMBL; X99694; E256957; -.
KW   PLASMID.
SQ   SEQUENCE   112 AA;  11986 MW;  7C5B0746 CRC32;


  Query Match            51.4%;  Score 38;  DB 9;  Length 112;
  Best Local Similarity  54.5%;  Pred. No. 23;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 HCDSSECKSSP 12
          ||:|: | :||
Db     99 HCESGSCTTSP 109


RESULT  10
Q25464
ID   Q25464      PRELIMINARY;      PRT;   473 AA.
AC   Q25464;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR.
OS   MYTILUS GALLOPROVINCIALIS.
OC   EUKARYOTA; METAZOA; MOLLUSCA; BIVALVIA; PTERIOMORPHIA; MYTILOIDA;
OC   MYTILIDAE; MYTILUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FOOT;
RX   MEDLINE; 95204464.
RA   INOUE K., TAKEUCHI Y., MIKI D., ODO S.;
RT   "Mussel adhesive plaque protein gene is a novel member of epidermal
RT   growth factor-like gene family.";
RL   J. BIOL. CHEM. 270:6698-6701(1995).
DR   EMBL; D43794; G602768; -.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PFAM; PF00008; EGF; 11.
KW   SIGNAL; MATRIX PROTEIN; GLYCOPROTEIN.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    473       ADHESIVE PLAQUE MATRIX PROTEIN.
```

```
SQ   SEQUENCE   473 AA;  51772 MW;  E6E2CD33 CRC32;


  Query Match             55.4%;  Score 41;  DB 3;  Length 473;
  Best Local Similarity   41.7%;  Pred. No. 26;
  Matches    5;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRC 14
           | :::||: :||
Db      86 CKPNQCKNKSRC 97


RESULT  11
Q16609
ID   Q16609      PRELIMINARY;      PRT;   132 AA.
AC   Q16609;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   (APOARGC).
GN   APOARGC.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 95268939.
RA   BYRNE C.D., SCHWARTZ K., LAWN R.M.;
RT   "Loss of a splice donor site at a 'skipped exon' in a gene homologous
RT   to apolipoprotein(a) leads to an mRNA encoding a protein consisting of
RT   a single kringle domain.";
RL   ARTERIOSCLER. THROMB. VASC. BIOL. 15:65-70(1995).
DR   EMBL; U19518; G642946; -.
DR   EMBL; U19517; G642944; -.
DR   PFAM; PF00051; kringle; 1.
SQ   SEQUENCE   132 AA;  14886 MW;  8E9987A2 CRC32;


  Query Match             51.4%;  Score 38;  DB 2;  Length 132;
  Best Local Similarity   33.3%;  Pred. No. 26;
  Matches    4;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRC 14
           | :::| ::|:|
Db      77 CRNPDCSAGPWC 88


RESULT  12
Q92070
ID   Q92070      PRELIMINARY;      PRT;   772 AA.
AC   Q92070;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   INTEGRIN BETA 2 SUBUNIT (CD18) PRECURSOR (CD18).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LEG-HORN; TISSUE=SPLEEN;
RX   MEDLINE; 94194252.
RA   BILSLAND C.A., SPRINGER T.A.;
RT   "Cloning and expression of the chicken CD18 cDNA.";
RL   J. LEUKOC. BIOL. 55:501-506(1994).
CC   -!- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE
CC       GLYCOPROTEINS THAT MEDIATE CELL TO CELL & CELL TO MATRIX ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- SIMILARITY: WITH OTHER BETA CHAINS FROM THE INTEGRIN FAMILY OF
CC       CELL-SURFACE RECEPTOR.
DR   EMBL; X71786; G297567; -.
DR   PROSITE; PS00243; INTEGRIN_BETA; 2.
DR   PFAM; PF00362; integrin_B; 1.
KW   SIGNAL; INTEGRIN; CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25    772       POTENTIAL.
SQ   SEQUENCE   772 AA;  85368 MW;  474AB87C CRC32;


  Query Match             56.8%;  Score 42;  DB 12;  Length 772;
  Best Local Similarity   50.0%;  Pred. No. 27;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 HCDSSECKSSPR 13
           :|| :||| :|:
Db     557 RCDCGECKCTPK 568


RESULT  13
O42347
ID   O42347      PRELIMINARY;      PRT;  1212 AA.
AC   O42347;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   C-SERRATE-2 (FRAGMENT).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97184054.
RA   HAYASHI H., MOCHII M., KODAMA R., HAMADA Y., MIZUNO N., EGUCHI G.,
RA   TACHI C.;
RT   "Isolation of a novel chick homolog of Serrate and its coexpression
RT   with C-Notch-1 in chick development.";
RL   INT. J. DEV. BIOL. 40:1089-1096(1996).
DR   EMBL; D87558; D1022568; -.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PFAM; PF00008; EGF; 14.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
FT   NON_TER      1      1
SQ   SEQUENCE  1212 AA;  134188 MW;  0ECF076C CRC32;


  Query Match             58.1%;  Score 43;  DB 12;  Length 1212;
  Best Local Similarity   46.2%;  Pred. No. 27;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRCK 15
           |:|: | :: |||
Db     472 CESNPCQNGGRCK 484


RESULT  14
Q40630
ID   Q40630      PRELIMINARY;      PRT;   218 AA.
AC   Q40630;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   OSMOTIN (FRAGMENT).
GN   14B.
OS   ORYZA SATIVA (RICE).
OC   EUKARYOTA; VIRIDIPLANTAE; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;
OC   TRACHEOPHYTA; EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA;
OC   POALES; POACEAE; ORYZA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CALLUS;
RA   MCGONIGLE B., LAI L.B., NELSON T.;
RL   PLANT PHYSIOL. 111:1354-1354(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
```

RC TISSUE=CALLUS;
RA MEDINA J., DAVIS J., QUATRANO R.S.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; L76377; G1196835; -.
DR PFAM; PF00314; thaumatin; 1.
FT NON_TER 1 1
SQ SEQUENCE 218 AA; 22430 MW; C2C0F7D0 CRC32;

Query Match 52.7%; Score 39; DB 8; Length 218;
Best Local Similarity 30.8%; Pred. No. 28;
Matches 4; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Qy 2 HCDSSECKSSPRC 14
:|::::| :: ||
Db * 81 QCQTGDCGGTLRC 93

RESULT 15
O18190
ID O18190 PRELIMINARY; PRT; 547 AA.
AC Q18190;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE W09G3.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z82080; E1188499; -.
SQ SEQUENCE 547 AA; 61992 MW; 580C12B4 CRC32;

Query Match 55.4%; Score 41; DB 3; Length 547;
Best Local Similarity 57.1%; Pred. No. 29;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 HCDSSECKSSPRCK 15
||| :| : ||||
Db 516 HCDFAEISNFPRCK 529

Search completed: May 3, 1999, 17:54:24
Job time: 644 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 21:34:32 ; Search time 854.75 Seconds
(without alignments)
5319.758 Million cell updates/sec

Title: US-09-037-460-1
Perfect score: 1271
Sequence: 1 CTGCTTCCCACCAGCAAAGA.....GTTNCCTNTNTCTNAAGGAC 1271

Scoring table: IDENTITY_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl1:*
9: gb_pl2:*
10: gb_pr1:*
11: gb_pr2:*
12: gb_pr3:*
13: gb_ro:*
14: gb_st:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: gb_htg:*
19: em_ba:*
20: em_fun:*
21: em_hum1:*
22: em_hum2:*
23: em_in:*
24: em_om:*
25: em_or:*
26: em_ov:*
27: em_pat:*
28: em_ph:*
29: em_pl:*
30: em_ro:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: em_htg:*
35: em_sts:*
36: gb_ba1:*
37: gb_ba2:*
38: gb_pl1:*
39: gb_pl2:*
40: gb_pr1:*
41: gb_pr2:*
42: gb_pr3:*
43: gb_sts:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1264.8	99.5	1271	6	AR004664	AR004664 Sequence
2	1204.8	94.8	2006	10	HSRNAESM1	X89426 H.sapiens m
3	1204.8	94.8	2006	40	HSRNAESM1	X89426 H.sapiens m
4	440	34.6	2009	13	RNU80818	U80818 Rattus norv
c 5	62.4	4.9	7218	6	I66494	I66494 Sequence 14
c 6	45	3.5	103911	11	HUAC002565	AC002565 Human Chr
c 7	45	3.5	103911	41	HUAC002565	AC002565 Human Chr
8	43.6	3.4	138187	13	MMMMH461	AF027865 Mus muscu
9	42.4	3.3	175817	18	AC005507	AC005507 *** SEQUE
c 10	42	3.3	10780	3	CEMYO2	X08066 Caenorhabdi
c 11	41.4	3.3	155644	18	AC006101	AC006101 *** SEQUE
c 12	41	3.2	99689	9	AC004557	AC004557 Genomic s
c 13	41	3.2	99689	39	AC004557	AC004557 Genomic s
c 14	41	3.2	111390	8	AC000348	AC000348 Genomic s
c 15	41	3.2	111390	38	AC000348	AC000348 Genomic s
c 16	40.8	3.2	13437	3	AE001407	AE001407 Plasmodiu
c 17	40.4	3.2	6104	3	PFSC04016	AL010228 Plasmodiu
c 18	40.4	3.2	180105	18	AC004688	AC004688 *** SEQUE
19	39.8	3.1	39873	18	CEW03A8	Z82066 *** SEQUENC
c 20	39.8	3.1	39923	3	CEY39G8C	AL032634 Caenorhab
21	39.8	3.1	110000	18	CEY39G8_0	Z92851 Caenorhabdi
c 22	39.8	3.1	110000	18	CEY39G8_2	Continuation (3 of
c 23	39.8	3.1	311976	18	CEY48B6	AL021151 Caenorhab
c 24	39.4	3.1	5850	8	DDIDDP2	M55298 Dictyosteli
c 25	39.4	3.1	5850	38	DDIDDP2	M55298 Dictyosteli
c 26	39.2	3.1	1191	3	AF044870	AF044870 Zerene eu
27	39.4	3.1	89817	12	AC005876	AC005876 Homo sapi
28	39.4	3.1	89817	42	AC005876	AC005876 Homo sapi
29	39.4	3.1	197783	18	AC004153	AC004153 *** SEQUE
c 30	39.2	3.1	33879	3	CELT05B4	AF016445 Caenorhab
31	39.2	3.1	105576	10	HS106I20B	Z81369 Human DNA s
32	39.2	3.1	105576	40	HS106I20B	Z81369 Human DNA s
c 33	39	3.1	16422	3	AE001421	AE001421 Plasmodiu
34	38.8	3.1	2118	3	PLRNASSU	X13706 Plasmodium
35	38.8	3.1	6970	3	PFSC03064	AL009014 Plasmodiu
36	38.8	3.1	19014	3	PFSC03009	AL008989 Plasmodiu
37	39	3.1	177503	12	AC005297	AC005297 Homo sapi
38	39	3.1	177503	42	AC005297	AC005297 Homo sapi
c 39	38.6	3.0	2993	3	PFSC04014	AL010217 Plasmodiu
40	38.8	3.1	27975	3	CET14G8	Z67884 Caenorhabdi
c 41	38.8	3.1	28871	3	CEC47B2	Z99709 Caenorhabdi
42	38.4	3.0	420	3	PMU78741	U78741 Plasmodium
c 43	38.6	3.0	3948	9	MEAJ3281	AJ223281 Manihot e
c 44	38.6	3.0	3948	39	MEAJ3281	AJ223281 Manihot e
c 45	38.8	3.1	192126	12	DJ270M14	AF107885 Homo sapi

ALIGNMENTS

RESULT 1
AR004664
LOCUS AR004664 1271 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5747280.
ACCESSION AR004664
NID g3965543
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1271)
AUTHORS Hastings,G.A. and Rosen,C.A.
TITLE Human vascular IBP-like growth factor
JOURNAL Patent: US 5747280-A 1 05-MAY-1998;
FEATURES Location/Qualifiers
source 1. .1271
/organism="unknown"
BASE COUNT 361 a 246 c 336 g 320 t 8 others
ORIGIN

Query Match 99.5%; Score 1264.8; DB 6; Length 1271;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60

Qy     61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120

Qy    121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180

Qy    181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240

Qy    241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300

Qy    301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360

Qy    361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420

Qy    421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480

Qy    481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540

Qy    541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600

Qy    601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAA 660

Qy    661 CACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAA 720

Qy    721 GACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780

Qy    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840

Qy    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900

Qy    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960

Qy    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020

Qy   1021 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080
```

```
Qy   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140

Qy   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200

Qy   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260

Qy   1261 NTCTNAAGGAC 1271
          ||||||||||||
Db   1261 NTCTNAAGGAC 1271
```

```
RESULT   2
HSRNAESM1
LOCUS       HSRNAESM1    2006 bp    RNA             PRI       07-OCT-1996
DEFINITION  H.sapiens mRNA for ESM-1 protein.
ACCESSION   X89426
NID         g1150418
KEYWORDS    ESM-1 protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2006)
  AUTHORS   Lassalle,P., Molet,S., Janin,A., Heyden,J.V., Tavernier,J.,
            Fiers,W., Devos,R. and Tonnel,A.B.
  TITLE     ESM-1 is a novel human endothelial cell-specific molecule expressed
            in lung and regulated by cytokines
  JOURNAL   J. Biol. Chem. 271 (34), 20458-20464 (1996)
  MEDLINE   96355375
REFERENCE   2  (bases 1 to 2006)
  AUTHORS   Lassalle,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1995) P.M. Lassalle, INSERM, Unite 416, 1, bd du
            Prof. CALMETTE, LILLE 59019, FRANCE
FEATURES             Location/Qualifiers
     source          1. .2006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HUVEC"
                     /clone="A11.1"
     sig_peptide     56. .112
     CDS             56. .610
                     /codon_start=1
                     /product="ESM-1 secretory protein"
                     /db_xref="PID:e189266"
                     /db_xref="PID:g1150419"
                     /translation="MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK
                     RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDC
                     PYGTFGMDCRETCNCQSGICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGN
                     IVREEVVKENAAGSPVMRKWLNPR"
BASE COUNT      623 a    333 c    475 g    575 t
ORIGIN
```

```
  Query Match             94.8%;  Score 1204.8;  DB 10;  Length 2006;
  Best Local Similarity   98.5%;  Pred. No. 1.1e-290;
  Matches 1252;  Conservative    3;  Mismatches   10;  Indels    6;  Gaps    4;

Qy       4 CTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGA-GCAGCTGGGAAACATGAA 62
            ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       1 CTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGGCAGCTGGGAAACATGAA 60

Qy      63 GAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAA 122
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAA 120
```

```
Qy    123 TAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCG 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCG 180

Qy    183 CTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGG 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGG 240

Qy    243 AGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTGAG 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTGAG 300

Qy    303 GTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGACTG 362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGACTG 360

Qy    363 TCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATCTG 422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATCTG 420

Qy    423 TGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAGTC 482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAGTC 480

Qy    483 TTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATTGT 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATTGT 540

Qy    543 GAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTAAA 602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTAAA 600

Qy    603 TCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAACA 662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db    601 TCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGATTGTTCAACA 660

Qy    663 CACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAAGA 722
          |||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db    661 CACAGCCAACATTTTAGGAACTTTCTAGA-TATAGCATAAGTACATGTAATTTTTGAAGA 719

Qy    723 --CCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780
            | || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 TCCAAATTGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 779

Qy    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 839

Qy    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    840 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 899

Qy    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 959

Qy    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1019

Qy   1021 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1079

Qy   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140
          ||||||||||||||:|||||||||||||||||  ||||||||||||||||||||||||||
Db   1080 CTCTAGAATTTAATTGTGCTTTTTTTTTTTTTTT--GGCTTTGGGAAAAGTCAAAATAAA 1137

Qy   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1138 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1197

Qy   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260
          ||||||||||| |||||||||||:||||||||||||||||||  |||||||| ||| |
Db   1198 AACAGCTTTGAACTGAGAGCAATTTCAAAAGGCTGCTGATGTAGTTCCCGGGTTACCTGT 1257

Qy   1261 NTCTNAAGGAC 1271
          ||| ||||||
Db   1258 ATCTGAAGGAC 1268
```

```
RESULT   3
HSRNAESM1
LOCUS       HSRNAESM1    2006 bp    RNA             PRI       07-OCT-1996
DEFINITION  H.sapiens mRNA for ESM-1 protein.
ACCESSION   X89426
NID         g1150418
KEYWORDS    ESM-1 protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2006)
  AUTHORS   Lassalle,P., Molet,S., Janin,A., Heyden,J.V., Tavernier,J.,
            Fiers,W., Devos,R. and Tonnel,A.B.
  TITLE     ESM-1 is a novel human endothelial cell-specific molecule expressed
            in lung and regulated by cytokines
  JOURNAL   J. Biol. Chem. 271 (34), 20458-20464 (1996)
  MEDLINE   96355375
REFERENCE   2  (bases 1 to 2006)
  AUTHORS   Lassalle,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1995) P.M. Lassalle, INSERM, Unite 416, 1, bd du
            Prof. CALMETTE, LILLE 59019, FRANCE
FEATURES             Location/Qualifiers
     source          1. .2006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HUVEC"
                     /clone="A11.1"
     sig_peptide     56. .112
     CDS             56. .610
                     /codon_start=1
                     /product="ESM-1 secretory protein"
                     /db_xref="PID:e189266"
                     /db_xref="PID:g1150419"
                     /translation="MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK
                     RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDC
                     PYGTFGMDCRETCNCQSGICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGN
                     IVREEVVKENAAGSPVMRKWLNPR"
BASE COUNT      623 a    333 c    475 g    575 t
ORIGIN
```

```
  Query Match           94.8%;  Score 1204.8;  DB 40;  Length 2006;
  Best Local Similarity  98.5%;  Pred. No. 1.1e-290;
  Matches 1252;  Conservative    3;  Mismatches   10;  Indels    6;  Gaps    4;

Qy      4 CTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGA-GCAGCTGGGAAACATGAA 62
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      1 CTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGGCAGCTGGGAAACATGAA 60

Qy     63 GAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAA 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAA 120

Qy    123 TAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCG 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCG 180

Qy    183 CTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGG 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGG 240

Qy    243 AGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTGAG 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 AGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTGAG 300

Qy    303 GTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGACTG 362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGACTG 360

Qy    363 TCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATCTG 422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATCTG 420

Qy    423 TGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAGTC 482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAGTC 480

Qy    483 TTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATTGT 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATTGT 540

Qy    543 GAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTAAA 602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTAAA 600

Qy    603 TCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAACA 662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db    601 TCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGATTGTTCAACA 660

Qy    663 CACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAAGA 722
          |||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||
Db    661 CACAGCCAACATTTTAGGAACTTTCTAGA-TATAGCATAAGTACATGTAATTTTTGAAGA 719

Qy    723 --CCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780
            | || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 TCCAAATTGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 779

Qy    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 839

Qy    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    840 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 899

Qy    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 959

Qy    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1019

Qy   1021 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1079

Qy   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140
          ||||||||||||||:|||||||||||||||||  ||||||||||||||||||||||||||
Db   1080 CTCTAGAATTTAATTGTGCTTTTTTTTTTTTTTT--GGCTTTGGGAAAAGTCAAAATAAA 1137

Qy   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1138 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1197

Qy   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260
          |||||||||| ||||||||||:||||||||||||||||||  |||||||| ||| |
Db   1198 AACAGCTTTGAACTGAGAGCAATTTCAAAAGGCTGCTGATGTAGTTCCCGGGTTACCTGT 1257

Qy   1261 NTCTNAAGGAC 1271
           ||| ||||||
Db   1258 ATCTGAAGGAC 1268


RESULT   4
RNU80818
```

```
LOCUS       RNU80818     2009 bp    mRNA            ROD       23-DEC-1996
DEFINITION  Rattus norvegicus pineal specific PG25 mRNA, complete cds.
ACCESSION   U80818
NID         g1750197
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2009)
  AUTHORS   Wang,X., Brownstein,M.J. and Young,W.S.
  TITLE     PG25, a full-length pineal-specific cDNA, cloned by differential
            display PCR (DDPCR) and rapid amplification of cDNA ends (RACE)
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 2009)
  AUTHORS   Wang,X., Brownstein,M.J. and Young,W.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1996) LCMR, NIMH, 36 Convent Drive, MSC 40688,
            Bethesda, MD 20892-4068, USA
FEATURES             Location/Qualifiers
     source          1. .2009
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /tissue_type="pineal gland"
     CDS             51. .605
                     /note="similar to human lung endothelial cell-specific
                     ESM-1 gene, encoded by Genbank Accession Number X89426"
                     /codon_start=1
                     /product="PG25"
                     /db_xref="PID:g1750198"
                     /translation="MKSLLLLTTLLIPLHLGMAWSAKYAVDCPEHCDNTECRSSLRCK
                     RTVLDDCGCCQVCAAGPGETCYRTVSGMDGVKCGPGLKCHFYSEEDDFGDEFGVCKDC
                     PYGTFGMDCKETCNCQSGICDRVTGRCLDFPFFQYAAAKSPSRTSASQTERDAASGDG
                     NAVREEIGDRNAARPSVMKWLNPR"
BASE COUNT      576 a    346 c    503 g    584 t
ORIGIN
```

```
  Query Match           34.6%;  Score 440;  DB 13;  Length 2009;
  Best Local Similarity  67.1%;  Pred. No. 8.3e-100;
  Matches  814;  Conservative    1;  Mismatches  336;  Indels   63;  Gaps   11;

Qy      1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60
          |||||||| |||||||||||||||||| | || | | ||| |          || |||||
Db      2 CTGCTTCCCGCCAGCAAAGACCACGACGGCAGGGGCCAGCTG--------CTGAGACATG 53

Qy     61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120
          |||||| ||||||| |||||||||||||| | |||  |||||||     |||||||||
Db     54 AAGAGCCTCTTGCTACTGACCACGCTCCTGATTCCTCTGCACCTGGGCATGGCCTGGAGC 113

Qy    121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180
             || |||||||||| |||||  |||| |||||||  | ||||||||  |||||||| |
Db    114 GCCAAGTATGCGGTGGATTGCCCGGAACATTGTGACAATACTGAGTGCAGAAGCAGCCTG 173

Qy    181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240
          || |||||||||||||||| || ||||||||||||||| ||||||||||| ||||
Db    174 CGTTGCAAGAGGACAGTGCTGGATGACTGTGGCTGCTGCCAAGTGTGCGCTGCGGGGCCT 233

Qy    241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300
          || ||||| |||||||| |||||||||||||||||| ||  | |||||||| ||||| |||
Db    234 GGGGAAACCTGCTACCGTACAGTCTCAGGCATGGACGGTGTCAAGTGTGGTCCGGGACTG 293

Qy    301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360
          | ||||||   ||  |  | ||| |||  ||| ||||| |||||||| |||| ||||||
Db    294 AAGTGTCACTTTTACAGCGAGGAAGATGATTTCGGTGACGAGTTTGGTGTCTGTAAAGAC 353

Qy    361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420
          ||||||||||||| |||||||||||||  ||||| ||||| |||||||| |||||
Db    354 TGTCCCTACGGCACATTCGGGATGGATTGCAAGGAGACTTGCAATTGCCAGTCGGGCATA 413

Qy    421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480
```

```
Db    414 TGTGACAGGGTGACCGGGAGGTGTCTGGACTTCCCCTTCTTCCAGTATGCAGCAGCCAAG 473

Qy    481 TCTTCCAACAG---ATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAAT 537

Db    474 TCGCCCAGCAGGACCTCTGCCTCCCAAACAGAGCGCGATGCAGCCTCTGGAGATGGCAAT 533

Qy    538 ATTGTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGG 597

Db    534 GCTGTGAGAGAAGAGATTGGAGACCGGAACGCTGCCCGGCCCTCTGTAAT---GAAATGG 590

Qy    598 TTAAATCCACGCTGATCCCGGCTGTGATTT-CTGAGAGAAGGCTCTATTTTCGTGAYTGT 656

Db    591 TTAAATCCGCGCTGATCCTGGCTGAGATTTGCCAAGAGGAGGCTCCACTTTAGTGAC--- 648

Qy    657 TCAACACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTT 716

Db    648 -CAACTTTCAGCCAACATTCTAAGAACCGTCTAGAACAAAGACTGTGGACATGTAATTTA 706

Qy    717 TGAAGACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCC 776

Db    707 TGGAGATCAAGTGTGATTCA---------GAAAAAAAAAAATGTAGGCTATACACAACCC 757

Qy    777 ATAACATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGAT 836

Db    758 ATGAGACACATATGACTGAACAATT-TAGATATTTGTCAAATATTTAAATTCATGGTTTT 816

Qy    837 TTGTTAAATGTGTGTGTAT-----AGTAACACTGAAGAACTAAAAATGCAATTTAGGTAA 891

Db    817 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCATCACCAAGG 876

Qy    892 TCTTACATGGAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAG 951

Db    877 AACTTGATATGCAATTGAGATGGTCTTACTGTCAAGGAGTAATCCAATCAAAGCTGAGGA 936

Qy    952 TCAAATTAGTTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGA----CT 1007

Db    937 AACTGTGGGTTTGTTGACTTGCGTGTACGTTCTTATTGAGATACAGGCTGAAGAAGAGTT 996

Qy   1008 TAAGAGCAGGAGAAGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCT 1067

Db    997 TGGAGTAAGGGACTGTGGAGATGGTAGCGAAAGTAGGGGCTATCATATTTGGCTCCTCCT 1056

Qy   1068 TGGTAGGTAGCTTCTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAA 1127

Db   1057 TG-----------TAGTAGATTAAAAGTCATTATTTTGTTTTAATTGTGGG--------- 1097

Qy   1128 AAGTCAAAATAAAACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGA 1187

Db   1097 -----AAATAAACAAGTCCAGAAAAGCCTTGAAGGAAGTGAGATATTGGAAGCTTATGGG 1151

Qy   1188 AATTTGAGTAACAA 1201

Db   1152 AGTCTAAACAACTA 1165
```

```
RESULT   5
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
```

```
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

Query Match 4.9%; Score 62.4; DB 6; Length 7218;
Best Local Similarity 5.9%; Pred. No. 1.5e-05;
Matches 24; Conservative 225; Mismatches 161; Indels 0; Gaps 0;

```
Qy    689 AGATTATAGCATAAGGACATGTAATTTTTGAAGACCAAATGTGATGCATGGTGGATCCAG 748

Db   1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372

Qy    749 AAAACAAAAAGTAGGATACTTACAATCCATAACATCCATATGACTGAACACTTGTATGTG 808

Db   1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312

Qy    809 TTTGTTAAATATTCGAATGCATGTAGATTTGTTAAATGTGTGTGTATAGTAACACTGAAG 868

Db   1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252

Qy    869 AACTAAAAAATGCAATTTAGGTAATCTTACATGGAGACAGGTCAACCAAAGAGGGAGCTAG 928

Db   1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192

Qy    929 GCAAAGCTGAAGACCGCAGTGAGTCAAATTAGTTCTTTGACTTTGATGTACATTAATGTT 988

Db   1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132

Qy    989 GGGATATGGAATGAAGACTTAAGAGCAGGAGAAGATGGGGAGGGGGTGGGAGTGGGAAAT 1048

Db   1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072

Qy   1049 AAAATATTTAGCCCTTCCTTGGTAGGTAGCTTCTCTAGAATTTAATTRTG 1098

Db   1071 RRRRRATCGCAAGCTCCCTCGACCTGCAGCCAAGCTCGGAATTAATTCTG 1022
```

```
RESULT   6
HUAC002565/c
LOCUS       HUAC002565 103911 bp    DNA             PRI       17-MAR-1998
DEFINITION  Human Chromosome 16 BAC clone CIT987SK-A-598D4, complete sequence.
ACCESSION   AC002565
NID         g2896804
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103911)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Human Chromosome 16 BAC clone CIT987SK-A-598D4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 103911)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 103911)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 103911)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Feb 19, 1998 this sequence version replaced gi:2894636.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: mdadams@tigr.org.  The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
```

```
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge, http://gnomic.stanford.Edu/'
            chris/GENSCANW.html)searches of the complete sequence against a
            peptide database,and the Human gene Index database at TIGR
            (http://www.tigr.org/tdb/hgi/hgi.html). Genes without pepetide
            homolgy having spliced EST hits are termed 'Unknown gene product'.
            Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES              Location/Qualifiers
     source           1. .103911
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="16"
                      /map="16p12"
                      /clone="A-598D4"
BASE COUNT    29634 a  22640 c  22484 g  29153 t
ORIGIN


  Query Match             3.5%;  Score 45;  DB 11;  Length 103911;
  Best Local Similarity   52.9%;  Pred. No. 0.39;
  Matches  119;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy     673 ATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAAGACCAAATGTGA 732
           | ||||  || | |  | || |||  |  |  | || | |||||| || |  | |  |
Db   84881 AATTTAATAAATATGAATATAATAAAAATAATATATTTTATTTTTAAATATAATAAATAT 84822

Qy     733 TGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAACATCCATATGAC 792
                 ||    |||| ||| |  |  |||  ||  |   |||| ||  ||||
Db   84821 AAATATAAATATAATAAAAATAAATATAATAATATATAATAAATATAATATATATTTT 84762

Qy     793 TGAACACTTGTATGTGTTTGTTAA-ATATTCGAATGCATGTAGATTTGTTAAATGTGTGT 851
           | ||   |  ||| | |||  ||| |||||  |||  || || || | ||||||  | | |
Db   84761 TAAAATATAATATATTTTTAATAATATATTTTAATTTATTTAAATATTTTAAATATATTT 84702

Qy     852 GTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTA 896
           | |   || |    | |  |||| ||  | | |   |||   ||
Db   84701 ATTTTAAAATATAATAAATATAAATATATATTATTAATAAATATA 84657


RESULT   7
HUAC002565/c
LOCUS       HUAC002565 103911 bp    DNA             PRI       17-MAR-1998
DEFINITION  Human Chromosome 16 BAC clone CIT987SK-A-598D4, complete sequence.
ACCESSION   AC002565
NID         g2896804
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 103911)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Human Chromosome 16 BAC clone CIT987SK-A-598D4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 103911)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 103911)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 103911)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Feb 19, 1998 this sequence version replaced gi:2894636.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: mdadams@tigr.org.  The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge, http://gnomic.stanford.Edu/'
            chris/GENSCANW.html)searches of the complete sequence against a
            peptide database,and the Human gene Index database at TIGR
            (http://www.tigr.org/tdb/hgi/hgi.html). Genes without pepetide
            homolgy having spliced EST hits are termed 'Unknown gene product'.
            Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES              Location/Qualifiers
     source           1. .103911
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="16"
                      /map="16p12"
                      /clone="A-598D4"
BASE COUNT    29634 a  22640 c  22484 g  29153 t
ORIGIN


  Query Match             3.5%;  Score 45;  DB 41;  Length 103911;
  Best Local Similarity   52.9%;  Pred. No. 0.39;
  Matches  119;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy     673 ATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAAGACCAAATGTGA 732
           | ||||  || | |  | || |||  |  |  | || | |||||| || |  | |  |
Db   84881 AATTTAATAAATATGAATATAATAAAAATAATATATTTTATTTTTAAATATAATAAATAT 84822

Qy     733 TGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAACATCCATATGAC 792
                 ||    |||| ||| |  |  |||  ||  |   |||| ||  ||||
Db   84821 AAATATAAATATAATAAAAATAAATATAATAATATATAATAAATATAATATATATTTT 84762

Qy     793 TGAACACTTGTATGTGTTTGTTAA-ATATTCGAATGCATGTAGATTTGTTAAATGTGTGT 851
           | ||   |  ||| | |||  ||| |||||  |||  || || || | ||||||  | | |
Db   84761 TAAAATATAATATATTTTTAATAATATATTTTAATTTATTTAAATATTTTAAATATATTT 84702

Qy     852 GTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTA 896
           | |   || |    | |  |||| ||  | | |   |||   ||
Db   84701 ATTTTAAAATATAATAAATATAAATATATATTATTAATAAATATA 84657


RESULT   8
MMMMH461
LOCUS       MMMMH461   138187 bp    DNA             ROD       22-OCT-1997
DEFINITION  Mus musculus Major Histocompatibility Locus class II region.
ACCESSION   AF027865
NID         g2555188
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 138187)
  AUTHORS   Rowen,L., Qin,S., Ahearn,M.E., Loretz,C., Faust,J., Lasky,S.,
            Mahairas,G. and Hood,L.
  TITLE     Sequence of the mouse major histocompatibility locus class II
            region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138187)
  AUTHORS   Rowen,L., Qin,S., Ahearn,M., Loretz,C., Faust,J., Lasky,S.,
            Mahairas,G. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-1997) Department of Molecular Biotechnology, Box
            357730 University of Washington, Seattle, Washington 98195, USA
COMMENT     Sequencing methodology: high redundancy shotgun. Interspersed
            Repeats were identified with RepeatMasker (available from
            http://ftp.genome.washington.edu/RM/RepeatMasker.html) Simple
            sequence repeats were identified with sputnik (available from
            http://serac.mbt.washington.edu/~chrisa/software/sputnik.html).
```

```
FEATURES             Location/Qualifiers
     source          1. .138187
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /clone="BAC461G2"
                     /clone_lib="Genome Systems mouse BAC library"
     repeat_region   complement(5. .117)
                     /rpt_family="MTE"
     repeat_region   complement(148. .253)
                     /rpt_family="ORR1C"
     repeat_region   complement(314. .538)
                     /rpt_family="B3"
     repeat_region   complement(546. .642)
                     /rpt_family="B1-F"
     repeat_region   complement(1597. .1922)
                     /rpt_family="ORR1A1"
     repeat_region   2331. .2520
                     /rpt_family="B2"
     repeat_region   2531. .2690
                     /rpt_family="RSINE1"
     repeat_region   2799. .2919
                     /rpt_family="B1-F"
     repeat_region   3428. .3547
                     /rpt_family="RSINE1"
     repeat_region   3758. .3834
                     /rpt_family="(CA)n"
     repeat_region   3828. .3934
                     /rpt_family="(GA)n"
     repeat_region   3874. .3904
                     /rpt_family="(CA)n"
     repeat_region   3940. .4061
                     /rpt_family="(CAGA)n"
     repeat_region   4125. .4243
                     /rpt_family="(GA)n"
     unsure          4281
                     /note="G could be an A"
                     /replace="a"
     repeat_region   complement(4655. .4969)
                     /rpt_family="MTE"
     repeat_region   complement(4833. .5044)
                     /rpt_family="MTC"
     repeat_region   5154. .5254
                     /rpt_family="RMER5"
     repeat_region   5979. .6334
                     /rpt_family="MTD"
     repeat_region   6539. .6675
                     /rpt_family="B3"
     repeat_region   6638. .6673
                     /rpt_family="(CA)n"
     repeat_region   6673. .6733
                     /rpt_family="(GA)n"
     repeat_region   complement(6844. .6978)
                     /rpt_family="B1_MM"
     repeat_region   complement(7109. .7459)
                     /rpt_family="ORR1D"
     repeat_region   complement(7516. .7625)
                     /rpt_family="B3A"
     repeat_region   complement(8201. .8276)
                     /rpt_family="B3"
     repeat_region   complement(8278. .8377)
                     /rpt_family="(GAAA)n"
     repeat_region   complement(8299. .8376)
                     /rpt_family="(GAA)n"
     repeat_region   complement(8385. .8498)
                     /rpt_family="PB1D7"
     repeat_region   complement(8508. .8542)
                     /rpt_family="(GAAAA)n"
     repeat_region   9718. .10121
                     /rpt_family="RLTR11A"
     repeat_region   complement(10114. .10185)
                     /rpt_family="(CA)n"
     repeat_region   complement(10555. .10639)
                     /rpt_family="MTE"
     repeat_region   11179. .11326
                     /rpt_family="B3"
     repeat_region   11728. .11839
                     /rpt_family="MT2A"
     repeat_region   11967. .12165
                     /rpt_family="MTE"
     repeat_region   12166. .12231
                     /rpt_family="MLT2FB"
     repeat_region   12219. .12310
                     /rpt_family="B3"
     repeat_region   complement(13034. .13253)
                     /rpt_family="B3"
     repeat_region   complement(13272. .13421)
                     /rpt_family="B1_MM"
     repeat_region   13515. .13549
                     /rpt_family="(TGG)n"
     repeat_region   complement(13550. .13709)
                     /rpt_family="RSINE1"
     repeat_region   13842. .14192
                     /rpt_family="MTB"
     repeat_region   complement(14294. .14377)
                     /rpt_family="PB1D9"
     repeat_region   complement(15035. .15153)
                     /rpt_family="RSINE1"
     repeat_region   complement(16138. .16401)
                     /rpt_family="B4A"
     repeat_region   16431. .16637
                     /rpt_family="B3"
     repeat_region   complement(16862. .16918)
                     /rpt_family="(TAAA)n"
     gene            complement(17042. .22163)
                     /gene="H2 lmp2"
     CDS             complement(join(17078. .17205,17943. .18084,18469. .18598,
                     19136. .19267,20584. .20651,22104. .22163))
                     /gene="H2 lmp2"
                     /note="Intron-exon boundaries defined using the cDNA
                     sequence found in GenBank Accession Number U22919 and
                     confirmed by EST hits"
                     /codon_start=1
                     /product="20S proteasome subunit lmp2"
                     /db_xref="PID:g2555189"
                     /translation="MLRAGAPTAGSFRTEEVHTGTTIMAVEFDGGVVVGSDSRVSAGT
                     AVVNRVFDKLSPLHQHIFCALSGSAADAQAIADMAAYQLELHGLELEEPPLVLAAANV
                     VKNISYKYREDLLAHLIVAGWDQCEGGQVYGTMGGMLIRQPFTIGGSGSSYIYGYVDA
                     AYKPGMTPEECRRFTTDAITLAMNRDGSSGGVIYLVTITAAGVDHRVILGDELPKFYD
                     E"
     repeat_region   17623. .17822
                     /rpt_family="B3A"
     repeat_region   complement(18204. .18350)
                     /rpt_family="B1-F"
     repeat_region   complement(19398. .19442)
                     /rpt_family="(CA)n"
     repeat_region   19448. .19496
                     /rpt_family="ID2"
     repeat_region   19512. .19638
                     /rpt_family="B1-F"
     repeat_region   complement(19735. .19861)
                     /rpt_family="B3"
     repeat_region   19955. .20093
                     /rpt_family="B1_MM"
     unsure          20070. .20072
                     /gene="H2 lmp2"
                     /note="3 As probably correct -- could be 2 As"
                     /replace="aa"
     gene            22734. .31636
                     /gene="H2 TAP1"
     CDS             join(22734. .23259,24037. .24151,24289. .24419,25422. .25627,
                     26188. .26385,27696. .27824,27987. .28175,28347. .28520,
                     28743. .28905,29706. .29842,31430. .31636)
                     /gene="H2 TAP1"
                     /note="intron-exon boundaries defined in relation to cDNAs
                     found in GenBank Accession Numbers X59615 and U60019, and
```

```
                also EST AA270189"
                /codon_start=1
                /product="antigen processing transporter TAP1"
                /db_xref="PID:g2555195"
                /translation="MAAHVWLAAALLLLVDWLLLRPMLPGIFSLLVPEVPLLRVWVVG
                LSRWAILGLGVRGVLGVTAGAHGWLAALQPLVAALSLALPGLALFRELAAWGTLREGD
                SAGLLYWNSRPDAFAISYVAALPAAALWHKLGSLWAPSGNRDAGDMLCRMLGFLGPKK
                RRLYLVLVLLILSCLGEMAIPFFTGRITDWILQDKTVPSFTRNIWLMSILTIASTALE
                FASDGIYNITMGHMHGRVHREVFRAVLRQETGFFLKNPAGSITSRVTEDTANVCESIS
                GTLSLLLWYLGRALCLLVFMFWGSPYLTLVTLINLPLLFLLPKKLGKVHQSLAVKVQE
                SLAKSTQVALEALSAMPTVRSFANEEGEAQKFRQKLEEMKTLNKKEALAYVAEVWTTS
                VSGMLLKVGILYLGGQLVIRGAVSSGNLVSFVLYQLQFTQAVQVLLSLYPSMQKAVGS
                SEKIFEYLDRTPCSPLSGSLAPSNMKGLVEFQDVSFAYPNQPKVQVLQGLTFTLHPGT
                VTALVGPNGSGKSTVAALLQNLYQPTGGQLLLDGQCLVQYDHHYLHTQVAAVGQEPLL
                FGRSFRENIAYGLNRTPTMEEITAVAVESGAHDFISGFPQGYDTEVGETGNQLSGGQR
                QAVALARALIRKPLLLILDDATSALDAGNQLRVQRLLYESPKRASRTVLLITQQLSLA
                EQAHHILFLREGSVGEQGTHLQLMKRGGCYRAMVEALAAPAD"
repeat_region   complement(23500. .23602)
                /rpt_family="(GA)n"
repeat_region   23610. .23698
                /rpt_family="(CCCTC)n"
repeat_region   complement(23715. .23853)
                /rpt_family="B1-F"
repeat_region   24557. .24664
                /rpt_family="B1-F"
repeat_region   complement(25102. .25218)
                /rpt_family="B1_MM"
repeat_region   25820. .25996
                /rpt_family="B3"
repeat_region   26862. .26969
                /rpt_family="PB1D7"
repeat_region   26982. .27138
                /rpt_family="RSINE1"
repeat_region   27140. .27174
                /rpt_family="(CAAA)n"
repeat_region   27289. .27520
                /rpt_family="B3"
repeat_region   27289. .27449
                /rpt_family="B3A"
repeat_region   27439. .27485
                /rpt_family="(CAGA)n"
repeat_region   complement(28203. .28322)
                /rpt_family="(GAA)n"
repeat_region   complement(29076. .29199)
                /rpt_family="(CA)n"
repeat_region   complement(29108. .29233)
                /rpt_family="(CATA)n"
repeat_region   complement(29236. .29359)
                /rpt_family="(CATA)n"
repeat_region   complement(29366. .29485)
                /rpt_family="(CA)n"
repeat_region   complement(29488. .29611)
                /rpt_family="(CATA)n"
repeat_region   complement(29613. .29664)
                /rpt_family="(CA)n"
repeat_region   complement(29864. .29979)
                /rpt_family="B1_MM"
repeat_region   complement(30230. .30360)
                /rpt_family="MIR"
repeat_region   complement(30380. .30594)
                /rpt_family="B3"
repeat_region   30639. .30779
                /rpt_family="B3"
repeat_region   31040. .31205
                /rpt_family="Lx6"
mRNA            join(33248. .33406,34094. .34241,34432. .34543,34996. .35125,
                35566. .35770,36033. .36301)
                /product="lmp7"
gene            33248. .36288
                /gene="H2-lmp7"
CDS             join(33260. .33406,34094. .34241,34432. .34543,34996. .35125,
                35566. .35770,36033. .36121)
                /gene="H2-lmp7"
                /note="intron-exon boundaries defined in relation to cDNAs
                in GenBank Accession Numbers X64449 and U22031."
                /codon_start=1
                /product="20S proteosome subunit lmp7"
                /db_xref="PID:g2555193"
                /translation="MALLDLCGAARGQRPEWAALDAGSGGRSDPGHYSFSAQAPELAL
                PRGMQPTAFLRSFGGDQERNVQIEMAHGTTTLAFKFQHGVIVAVDSRATAGSYISSLR
                MNKVIEINPYLLGTMSGCAADCQYWERLLAKECRLYYLRNGERISVSAASKLLSNMML
                QYRGMGLSMGSMICGWDKKGPGLYYVDDNGTRLSGQMFSTGSGNTYAYGVMDSGYRQD
                LSPEEAYDLGRRAIAYATHRDNYSGGVVNMYHMKEDGWVKVESSDVSDLLYKYGEAAL
                "
repeat_region   34649. .34692
                /rpt_family="(TA)n"
repeat_region   complement(35232. .35433)
                /rpt_family="B2"
polyA_signal    36283. .36288
                /gene="H2-lmp7"
repeat_region   complement(38911. .39027)
                /rpt_family="(GGA)n"
repeat_region   complement(39141. .39207)
                /rpt_family="POLY_A"
repeat_region   complement(39936. .39975)
                /rpt_family="(CA)n"
gene            40150. .50948
                /gene="H2 TAP2"
CDS             join(40150. .40639,40725. .40839,43609. .43739,43945. .44150,
                46725. .46922,47085. .47213,48870. .49058,49178. .49351,
                49501. .49660,50189. .50325,50769. .50948)
                /gene="H2 TAP2"
                /note="intron-exon boundaries defined in relation to cDNAs
                in GenBank Accession Numbers M90459 and U60087"
                /codon_start=1
                /product="antigen processing transporter TAP2"
                /db_xref="PID:g2555194"
                /translation="MALSYLRPWVSLLLADMALLGLLQGSLGNLLPQGLPGLWIEGTL
                RLGVLWGLLKVGELLGLVGTLLPLLCLATPLFFSLRALVGGTASTSVVRVASASWGWL
                LAGYGAVALSWAVWAVLSPAGVQEKEPGQENRTLMKRLLKLSRPDLPFLIAAFFFLVV
                AVWGETLIPRYSGRVIDILGGDFDPDAFASAIFFMCLFSVGSSFSAGCRGGSFLFTMS
                RINLRIREQLFSSLLRQDLGFFQETKTGELNSRLSSDTSLMSRWLPFNANILLRSLVK
                VVGLYFFMLQVSPRLTFLSLLDLPLTIAAEKVYNPRHQAVLKEIQDAVAKAGQVVREA
                VGGLQTVRSFGAEEQEVSHYKEALERCRQLWWRRDLEKDVYLVIRRVMALGMQVLILN
                CGVQQILAGEVTRGGLLSFLLYQEEVGQYVRNLVYMYGDMLSNVGAAEKVFSYLDRKP
                NLPQPGILAPPWLEGRVEFQDVSFSYPRRPEKPVLQGLTFTLHPGTVTALVGPNGSGK
                STVAALLQNLYQPTGGQLLLDGEPLTEYDHHYLHRQVVLVGQEPVLFSGSVKDNIAYG
                LRDCEDAQVMAAAQAACADDFIGEMTNGINTEIGEKGGQLAVGQKQRLAIARALVRNP
                RVLILDEATSALDAQCEQALQNWRSQGDRTMLVIAHRLHTVQNADQVLVLKQGRLVEH
                DQLRDGQDVYAHLVQQRLEA"
repeat_region   complement(40947. .41075)
                /rpt_family="B2"
repeat_region   complement(41518. .41629)
                /rpt_family="RSINE1"
repeat_region   41684. .41986
                /rpt_family="Lx"
repeat_region   42012. .42092
                /rpt_family="(CAAAA)n"
repeat_region   42099. .42248
                /rpt_family="B1_MM"
repeat_region   42252. .42277
                /rpt_family="POLY_A"
repeat_region   42561. .42818
                /rpt_family="B4A"
repeat_region   complement(42789. .42899)
                /rpt_family="(CATG)n"
repeat_region   42976. .43031
                /rpt_family="(CA)n"
repeat_region   complement(43172. .43255)
                /rpt_family="B1_MM"
repeat_region   complement(43518. .43618)
                /rpt_family="(GGGA)n"
repeat_region   44229. .44347
                /rpt_family="(TAGA)n"
repeat_region   44261. .44370
                /rpt_family="(CAGA)n"
```

```
repeat_region   complement(44390. .44455)
                /rpt_family="L1MB6"
repeat_region   complement(44641. .44845)
                /rpt_family="B3"
repeat_region   complement(44885. .44925)
                /rpt_family="(CA)n"
repeat_region   complement(45024. .45162)
                /rpt_family="L1MA4A"
repeat_region   complement(45080. .45173)
                /rpt_family="L1MB6"
repeat_region   complement(45850. .45901)
                /rpt_family="L1M2_orf2"
repeat_region   45915. .46013
                /rpt_family="B3"
repeat_region   46021. .46082
                /rpt_family="(CA)n"
repeat_region   complement(46097. .46151)
                /rpt_family="L1M4_orf2"
repeat_region   complement(46162. .46367)
                /rpt_family="B3"
repeat_region   complement(46539. .46573)
                /rpt_family="B4"
repeat_region   46580. .46606
                /rpt_family="(CTCCC)n"
repeat_region   48568. .48676
                /rpt_family="PB1D7"
repeat_region   49706. .49779
                /rpt_family="B4"
repeat_region   49796. .49832
                /rpt_family="(CAAAA)n"
repeat_region   49860. .49965
                /rpt_family="(CA)n"
repeat_region   51181. .51396
                /rpt_family="MTE"
repeat_region   complement(51553. .51580)
                /rpt_family="(CATA)n"
repeat_region   complement(51583. .51711)
                /rpt_family="B1_MM"
repeat_region   complement(51742. .51814)
                /rpt_family="(CA)n"
repeat_region   complement(52081. .52411)
                /rpt_family="MTD"
repeat_region   52414. .52446
                /rpt_family="(CAGA)n"
repeat_region   complement(52483. .52814)
                /rpt_family="ORR1D"
repeat_region   complement(52859. .53087)
                /rpt_family="B3"
repeat_region   54071. .54128
                /rpt_family="B2"
repeat_region   54113. .54212
                /rpt_family="B3"
repeat_region   54216. .54553
                /rpt_family="IAPLTR1_MM"
repeat_region   54554. .54814
                /rpt_family="IAP"
repeat_region   54803. .55608
                /rpt_family="IAP"
repeat_region   55596. .58668
                /rpt_family="IAP"
repeat_region   58669. .59006
                /rpt_family="IAPLTR1_MM"
repeat_region   59037. .59147
                /rpt_family="RSINE1"
repeat_region   59188. .59246
                /rpt_family="(TAAAA)n"
repeat_region   complement(59352. .59413)
                /rpt_family="B3"
repeat_region   complement(59745. .59891)
                /rpt_family="RSINE1"
repeat_region   59898. .60199
                /rpt_family="MTA"
repeat_region   complement(60203. .60612)
                /rpt_family="RMER19"
repeat_region   60615. .60962
                /rpt_family="IAPLTR1_MM"
repeat_region   60971. .67441
                /rpt_family="IAP"
repeat_region   67442. .67789
                /rpt_family="IAPLTR1_MM"
repeat_region   complement(67799. .67955)
                /rpt_family="RMER19"
repeat_region   complement(68000. .68305)
                /rpt_family="L1MA7"
repeat_region   complement(68330. .68429)
                /rpt_family="PB1D9"
repeat_region   complement(68448. .68655)
                /rpt_family="B3A"
repeat_region   complement(68771. .68814)
                /rpt_family="RSINE1"
repeat_region   68960. .69002
                /rpt_family="(CAA)n"
repeat_region   complement(69281. .69395)
                /rpt_family="L1M2_orf2"
repeat_region   complement(69407. .69625)
                /rpt_family="B2"
repeat_region   70059. .70179
                /rpt_family="MTD"
repeat_region   70179. .70326
                /rpt_family="B2"
repeat_region   70342. .70396
                /rpt_family="B2"
repeat_region   70396. .70453
                /rpt_family="B2"
repeat_region   70445. .70625
                /rpt_family="MTC"
repeat_region   complement(71163. .71338)
                /rpt_family="RMER19"
repeat_region   complement(71352. .71486)
                /rpt_family="RMER19"
repeat_region   complement(71532. .71603)
                /rpt_family="RMER19"
repeat_region   complement(71848. .72013)
                /rpt_family="RMER19"
repeat_region   72015. .72208
                /rpt_family="B2"
repeat_region   complement(72512. .72586)
                /rpt_family="(CA)n"
repeat_region   72584. .72987
                /rpt_family="MTD"
repeat_region   complement(73050. .73102)
                /rpt_family="(TAA)n"
repeat_region   complement(73092. .73141)
                /rpt_family="PB1D10"
unsure          73739. .73748
                /note="possibly 11 Ts rather than 10 Ts"
                /replace="ttttttttttt"
mRNA            join(73821. .73911,75857. .76126,77286. .77567,78317. .78427,
                78737. .78765,78938. .79265)
                /gene="H2-IAbeta2"
                /product="IAbeta2"
gene            73821. .79265
                /gene="H2-IAbeta2"
CDS             join(73821. .73911,75857. .76126,77286. .77567,78317. .78427,
                78737. .78765,78938. .78970)
                /gene="H2-IAbeta2"
                /note="intron-exon boundaries defined in relation to EST
                AA175921 and the orthologous human gene HLA-DObeta,
                GenBank Accession Number X03066."
                /codon_start=1
                /product="IAbeta2 subunit"
                /db_xref="PID:g2555191"
                /translation="MGAGRAPWVVALLVNLMRLDSFMIEGRDSPENFVIQAKADCYFT
                NGTEKVHLLVRFIFNLEEYLHFDSDLGMFVALTELGEPDADQWNKRLDLLETSRAAVN
                MVCRQKYKLGAPFTVERNVPPEVTVYPERTPLLQQHNLLLCSVTGFYPGDISVKWFRN
                GQEERSGVMSTGLVRNGDWTFQTTVMLEMIPELGDIYSCLVEHPGLLRPVSVAWMAQS
```

```
                EYSWKKILSGAAVFLLGLIVFLVGVVIHLKAQKASVETQPGNEASRESLHSQP"
repeat_region   74035. .74186
                /rpt_family="B1_MM"
repeat_region   74298. .74335
                /rpt_family="(GGGA)n"
repeat_region   75252. .75335
                /rpt_family="RSINE1"
repeat_region   complement(76231. .76359)
                /rpt_family="RSINE1"
repeat_region   76377. .77037
                /rpt_family="RMER19"
repeat_region   77834. .78147
                /rpt_family="B4"
repeat_region   78126. .78185
                /rpt_family="(CA)n"
repeat_region   78508. .78576
                /rpt_family="(GGGA)n"
repeat_region   79940. .80256
                /rpt_family="CHARLIE1"
repeat_region   81100. .81169
                /rpt_family="(GA)n"
unsure          81168. .81178
                /note="Unclear region in a microsatellite -- reverse reads
                conflict with forward reads."
repeat_region   complement(81383. .81530)
                /rpt_family="B1_MM"
repeat_region   81588. .81673
                /rpt_family="RSINE1"
repeat_region   complement(82009. .82224)
                /rpt_family="B3"
repeat_region   complement(82359. .82504)
                /rpt_family="B1_MM"
repeat_region   complement(82426. .82525)
                /rpt_family="MT2B"
repeat_region   complement(82664. .82752)
                /rpt_family="PB1D9"
repeat_region   complement(83255. .83311)
                /rpt_family="L1M4_orf2"
repeat_region   complement(83540. .83599)
                /rpt_family="(GA)n"
repeat_region   complement(83605. .83648)
                /rpt_family="PB1D9"
repeat_region   complement(83607. .83631)
                /rpt_family="B1-F"
repeat_region   complement(83872. .84181)
                /rpt_family="L1M4_orf2"
repeat_region   84419. .84521
                /rpt_family="B1-F"
repeat_region   84532. .84625
                /rpt_family="(CA)n"
repeat_region   complement(85716. .86114)
                /rpt_family="MTC"
repeat_region   complement(86116. .86188)
                /rpt_family="ID2"
repeat_region   86232. .86342
                /rpt_family="B1_MM"
repeat_region   86929. .87064
                /rpt_family="B1_MM"
repeat_region   87055. .87085
                /rpt_family="(AAAAC)n"
repeat_region   87100. .87299
                /rpt_family="B3"
repeat_region   complement(87520. .87580)
                /rpt_family="L1M4_orf2"
repeat_region   complement(87583. .87609)
                /rpt_family="(TAAA)n"
repeat_region   87670. .88055
                /rpt_family="MTD"
repeat_region   88182. .88319
                /rpt_family="B1_MM"
repeat_region   89799. .89843
                /rpt_family="(GAAAA)n"
repeat_region   89934. .89979
                /rpt_family="PB1D7"
repeat_region   90233. .90519
                /rpt_family="L1M4_orf2"
repeat_region   complement(90520. .90617)
                /rpt_family="(TAGA)n"
repeat_region   90630. .91358
                /rpt_family="RMER17A"
repeat_region   90975. .91012
                /rpt_family="(CT)n"
repeat_region   complement(94032. .94175)
                /rpt_family="B1-F"
gene            98161. .103957
                /gene="H2-IAbeta"
CDS             join(98161. .98254,99610. .99882,102185. .102466,
                102772. .102882,103310. .103333,103944. .103957)
                /gene="H2-IAbeta"
                /note="intron-exon boundaries defined in relation to cDNA
                in GenBank Accession Number M66213"
                /codon_start=1
                /product="IAbeta subunit"
                /db_xref="PID:g2555192"
                /translation="MALQIPSLLLSAAVVVLMVLSSPGTEGGDSERHFVYQFMGECYF
                TNGTQRIRYVTRYIYNREEYVRYDSDVGEHRAVTELGRPDAEYWNSQPEILERTRAEL
                DTVCRHNYEGPETHTSLRRLEQPNVVISLSRTEALNHHNTLVCSVTDFYPAKIKVRWF
                RNGQEETVGVSSTQLIRNGDWTFQVLVMLEMTPRRGEVYTCHVEHPSLKSPITVEWRA
                QSESAWSKMLSGIGGCVLGVIFLGLGLFIRHRSQKGPRGPPPAGLLQ"
repeat_region   complement(100722. .100843)
                /rpt_family="(GGAA)n"
repeat_region   complement(100781. .100898)
                /rpt_family="(GA)n"
repeat_region   complement(100810. .100925)
                /rpt_family="(GAAA)n"
repeat_region   complement(100937. .101084)
                /rpt_family="B1_MM"
repeat_region   complement(101091. .101567)
                /rpt_family="L1_MM"
repeat_region   complement(101879. .101971)
                /rpt_family="B1-F"
repeat_region   complement(104198. .104650)
                /rpt_family="RLTR11B"
repeat_region   complement(105588. .105779)
                /rpt_family="B2"
repeat_region   106318. .106408
                /rpt_family="(CA)n"
repeat_region   106408. .106455
                /rpt_family="(GA)n"
repeat_region   complement(106458. .106521)
                /rpt_family="B3"
repeat_region   106513. .107090
                /rpt_family="RMER19"
repeat_region   107021. .107129
                /rpt_family="RMER19"
repeat_region   complement(107131. .107220)
                /rpt_family="B3"
repeat_region   complement(108175. .108270)
                /rpt_family="RMER19"
repeat_region   108611. .108775
                /rpt_family="MYSERV"
repeat_region   108977. .109122
                /rpt_family="B1_MM"
repeat_region   complement(109223. .109498)
                /rpt_family="ORR1A2"
repeat_region   complement(109511. .109920)
                /rpt_family="Lx"
repeat_region   complement(109921. .109979)
                /rpt_family="ORR1A3"
repeat_region   109984. .110961
                /rpt_family="MYSERV"
unsure          110389
                /note="low quality region"
                /replace="c"
repeat_region   110723. .110766
                /rpt_family="(CAT)n"
```

```
repeat_region   110988. .111483
                /rpt_family="RLTR11A"
repeat_region   111492. .111926
                /rpt_family="MYSERV"
repeat_region   complement(111971. .112092)
                /rpt_family="RMER6A"
repeat_region   complement(112081. .112376)
                /rpt_family="ORR1A3"
repeat_region   complement(112377. .112688)
                /rpt_family="RMER6A"
repeat_region   complement(112607. .112802)
                /rpt_family="RMER6A"
repeat_region   complement(112890. .113055)
                /rpt_family="RMER12"
repeat_region   113056. .113394
                /rpt_family="RMER19"
repeat_region   113608. .113958
                /rpt_family="RMER19"
repeat_region   113987. .114402
                /rpt_family="BGLII"
repeat_region   114458. .114540
                /rpt_family="(CATA)n"
repeat_region   complement(114546. .115184)
                /rpt_family="RMER12"
repeat_region   complement(116802. .117079)
                /rpt_family="MTE"
repeat_region   complement(117275. .117353)
                /rpt_family="PB1D7"
mRNA            complement(join(117617. .117869,118001. .118175,
                118421. .118702,119134. .119382,122528. .122612))
                /product="IAalpha"
polyA_signal    complement(117617. .117622)
                /gene="H2-IAalpha"
gene            complement(117617. .122528)
                /gene="H2-IAalpha"
CDS             complement(join(118021. .118175,118421. .118702,
                119134. .119382,122528. .122612))
                /note="exons defined in relation to cDNA in database,
                GenBank Accession Number K01922 and EST AA260586"
                /codon_start=1
                /product="IAalpha subunit"
                /db_xref="PID:g2555190"
                /translation="MPRSRALILGVLALTTMLSLCGGEDDIEADHVGTYGISVYQSPG
                DIGQYTFEFDGDELFYVDLDKKETVWMLPEFGQLASFDPQGGLQNIAVVKHNLGVLTK
                RSNSTPATNEAPQATVFPKSPVLLGQPNTLICFVDNIFPPVINITWLRNSKSVADGVY
                ETSFFVNRDYSFHKLSYLTFIPSDDDIYDCKVEHWGLEEPVLKHWEPEIPAPMSELTE
                TVVCALGLSVGLVGIVVGTIFIIQGLRSGGTSRHPGPL"
repeat_region   119667. .119709
                /rpt_family="(CA)n"
repeat_region   119709. .119734
                /rpt_family="(GA)n"
repeat_region   complement(121877. .121980)
                /rpt_family="PB1D10"
repeat_region   123899. .124078
                /rpt_family="MLT2E"
repeat_region   123936. .124134
                /rpt_family="MLT2CA"
repeat_region   124336. .124374
                /rpt_family="(CA)n"
repeat_region   124375. .124473
                /rpt_family="RMER15"
repeat_region   124561. .124774
                /rpt_family="B3"
repeat_region   124871. .125020
                /rpt_family="B3"
repeat_region   complement(125171. .125368)
                /rpt_family="B3A"
repeat_region   125550. .125631
                /rpt_family="L1ME2"
repeat_region   complement(125925. .125982)
                /rpt_family="PB1D10"
repeat_region   125989. .126179
                /rpt_family="URR1A"
repeat_region   127683. .127870
                /rpt_family="B3"
repeat_region   127841. .127890
                /rpt_family="(AC)n"
repeat_region   complement(128282. .128349)
                /rpt_family="(CAT)n"
repeat_region   129374. .129525
                /rpt_family="RSINE1"
repeat_region   129526. .129547
                /rpt_family="(CA)n"
repeat_region   complement(130250. .130455)
                /rpt_family="B3"
repeat_region   131334. .131397
                /rpt_family="(GGAA)n"
repeat_region   complement(131734. .131856)
                /rpt_family="B1-F"
repeat_region   complement(131764. .131856)
                /rpt_family="PB1D10"
repeat_region   complement(131797. .131934)
                /rpt_family="B4A"
repeat_region   complement(132249. .132371)
                /rpt_family="B1_MM"
repeat_region   133004. .133251
                /rpt_family="Lx9"
repeat_region   133785. .133903
                /rpt_family="(CAGA)n"
repeat_region   complement(134046. .134115)
                /rpt_family="PB1D7"
repeat_region   complement(134513. .134706)
                /rpt_family="B3"
repeat_region   complement(134802. .135034)
                /rpt_family="B3"
repeat_region   134827. .134906
                /rpt_family="(GT)n"
repeat_region   complement(134858. .135034)
                /rpt_family="B3A"
repeat_region   complement(136502. .136550)
                /rpt_family="(TAAA)n"
repeat_region   complement(136784. .136887)
                /rpt_family="B1-F"
repeat_region   136991. .137150
                /rpt_family="B3"
repeat_region   complement(137151. .137200)
                /rpt_family="(GA)n"
repeat_region   137201. .137276
                /rpt_family="(CATA)n"
repeat_region   137201. .137254
                /rpt_family="(CA)n"
BASE COUNT    35324 a  32541 c  32732 g  37590 t
ORIGIN


  Query Match             3.4%;  Score 43.6;  DB 13;  Length 138187;
  Best Local Similarity   54.5%;  Pred. No. 0.89;
  Matches   85;  Conservative    1;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    1009 AAGAGCAGGAGAAGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTT 1068
           | | ||||||||| ||||||   ||| |||||         | ||  |  | | |  | |
Db   49690 AGGGGCAGGAGAAAATGGGGCTAGGGAGGGAGCTCATGTAACAAAGTGCATCGTTGCAAG 49749

Qy    1069 GGTAGGTAGCTTCTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAA 1128
           |  |   ||  |  | ||     |  :|  ||||||||||||||||          |||
Db   49750 CATGAGGGCCTGCGTTTGATCGCTAGAATCTTTTTTTTTTTTTTTTAAACGAAAACCAAA 49809

Qy    1129 AGTCAAAATAAAACAACCAGAAAACCCCTGAAGGAA 1164
           |  ||||| |||||||    ||||||     |||||
Db   49810 AAGCAAAACAAAACAAAACAAAAACCAGACTAGGAA 49845


RESULT   9
AC005507
LOCUS       AC005507   175817 bp    DNA             HTG       02-DEC-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Plasmodium falciparum 3D7 chromosome
```

```
            12 PFYACB8-628 genomic sequence; HTGS phase 1, 23 unordered pieces.
ACCESSION   AC005507
NID         g3947406
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 175817)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175817)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Dec 2, 1998 this sequence version replaced gi:3785954.
            ***                                                          ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                          ***
            * This sequence is unfinished. It consists of 23 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *       1     4252: contig of 4252 bp in length
            *    4253     4302: gap of unknown length
            *    4303    13546: contig of 9244 bp in length
            *   13547    13596: gap of unknown length
            *   13597    30165: contig of 16569 bp in length
            *   30166    30215: gap of unknown length
            *   30216    40515: contig of 10300 bp in length
            *   40516    40565: gap of unknown length
            *   40566    48776: contig of 8211 bp in length
            *   48777    48826: gap of unknown length
            *   48827    54664: contig of 5838 bp in length
            *   54665    54714: gap of unknown length
            *   54715    62601: contig of 7887 bp in length
            *   62602    62651: gap of unknown length
            *   62652    65434: contig of 2783 bp in length
            *   65435    65484: gap of unknown length
            *   65485    90084: contig of 24600 bp in length
            *   90085    90134: gap of unknown length
            *   90135   109461: contig of 19327 bp in length
            *  109462   109511: gap of unknown length
            *  109512   120331: contig of 10820 bp in length
            *  120332   120381: gap of unknown length
            *  120382   129750: contig of 9369 bp in length
            *  129751   129800: gap of unknown length
            *  129801   136222: contig of 6422 bp in length
            *  136223   136272: gap of unknown length
            *  136273   142626: contig of 6354 bp in length
            *  142627   142676: gap of unknown length
            *  142677   147947: contig of 5271 bp in length
            *  147948   147997: gap of unknown length
            *  147998   153135: contig of 5138 bp in length
            *  153136   153185: gap of unknown length
            *  153186   157105: contig of 3920 bp in length
            *  157106   157155: gap of unknown length
            *  157156   160654: contig of 3499 bp in length
            *  160655   160704: gap of unknown length
            *  160705   163855: contig of 3151 bp in length
            *  163856   163905: gap of unknown length
            *  163906   167011: contig of 3106 bp in length
            *  167012   167061: gap of unknown length
            *  167062   170056: contig of 2995 bp in length
            *  170057   170106: gap of unknown length
            *  170107   173045: contig of 2939 bp in length
            *  173046   173095: gap of unknown length
            *  173096   175817: contig of 2722 bp in length.

FEATURES             Location/Qualifiers
     source          1. .175817
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
BASE COUNT    68768 a  18961 c  19291 g  67639 t   1158 others
ORIGIN


  Query Match             3.3%;  Score 42.4;  DB 18;  Length 175817;
  Best Local Similarity   54.5%;  Pred. No. 1.8;
  Matches   85;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     744 TCCAGAAAACAAAAAGTAGGATACTTACAATCCATAACATCCATATGACTGAACACTTGT 803
           ||   || |  | | |||  |||  |||| |  |||| ||  |||||| |||    |
Db   29180 TCACAAATAATACATGTACAATATATACAGTAAATAATATAAATATGAATGATGCATATA 29239

Qy     804 ATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGTTAAATGTGTGTGTATAGTAACAC 863
           | |  | | ||||  |   ||||  || |  |  |  |||||   | | |  |||||
Db   29240 AAGAATATTTTAATAAAATGAATTTATCTTTAACTTATAAATTAATCTTTCCAGTAATAT 29299

Qy     864 TGAAGAACTAAAAATGCAATTTAGGTAATCTTACAT 899
           | || || ||    |  ||| |   | || | | ||
Db   29300 TTAATAATTATTTTTTAAATCTCCTTCATATGAAAT 29335


RESULT  10
CEMYO2/c
LOCUS       CEMYO2      10780 bp    DNA             INV       29-MAR-1993
DEFINITION  Caenorhabditis elegans myo-2 gene for myosin heavy chain 2 (MHC-C).
ACCESSION   X08066
NID         g6787
KEYWORDS    myo-2 gene; myosin; myosin heavy chain.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
            Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
            Rhabditidae; Caenorhabditis.
REFERENCE   1  (bases 1 to 10780)
  AUTHORS   Karn,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1988) Karn J., MRC Laboratory of Molecular
            Biology, Hills Road, Cambridge, United Kingdom
REFERENCE   2  (bases 1 to 10780)
  AUTHORS   Dibb,N.J., Maruyama,I., Krause,M. and Karn,J.
  TITLE     Sequence analysis of myosin heavy chain gene family from
            caenorhabditis elegans
  JOURNAL   J. Mol. Biol. (1988) In press
COMMENT     SG 25-39 and lambda sup3-2;
            see J01050 for overlapping sequence
            Entries M37233, M37235 and M37236 are all fragments of this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .10780
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /tissue_type="muscle (pharynx)"
                     /clone="lambda"
     exon            <1727. .1789
                     /number=1
     CDS             join(1727. .1789,1900. .2028,2084. .2224,2274. .2748,
                     2804. .2985,3034. .3193,3454. .4745,4798. .7334,7381. .7531,
                     7580. .7870,8060. .8341,8388. .8528)
                     /codon_start=1
                     /product="myosin heavy chain 2"
                     /db_xref="PID:g295767"
                     /db_xref="SWISS-PROT:P12845"
                     /translation="MDYENDPGWKYLRRSREEMLQDQSRAYDSKKNVWIPDSEDGYIE
                     GVITKTAGDNVTVSIGQGAEKTVKKDVVQEMNPPKFEKTEDMSNLTFLNDASVLYNLK
                     ARYAAMLIYTYSGLFCVVINPYKRLPIYTESVARMFMGKRRTEMPPHLFAVSDEAYRN
                     MLQNHENQSMLITGESGAGKTENTKKVISYFAAVGAAQQETFGAKKAATEEDKNKKKV
                     TLEDQIVQTNPVLEAFGNAKTVRNNNSSRFGKFIRIHFSKQGRVASCDIEHYLLEKSR
```

VIRQAPGERSYHIFYQVFSDYLPNLKKDLLLNKPVKDYWFIAQAELIIDGINDKEEHQ
LTDEAFDILKFTPTEKMECYRLVAAMMHMGNMKFKQRPREEQAEPDGTDDAERAAKCF
GIDSEEFLKALTRPRVKVGNEWVNKGQKIEQVNWAVGAMAKGLYSRIFNWLVKKCNQT
LDQKGISRDHFIGVLDIAGFEIFDFNSFEQLWINFVNEKLQQFFNHHMFVLEQEEYAR
EGIQWTFIDFGLDLQACIELIEKPLGIIAMLDEECIVPKATDLTLAQKLIDQHLGKHP
NFEKPKPPKGKQAEAHFAMRHYAGTVRYNCLNWLEKNKDPLNDTVVTVMKASKEHALI
VEVWQDYTTQEEAAAAAAKGTAGAKKKGKSGSFMTVSMLYRESLNKLMTMLHSTHPHF
IRCIIPNEKKASGVIDAGLVLNQLTCNGVLEGIRICRKGFPNRTLHPDFVQRYALLAA
DESIIGKTDAKKGSALMLARLVKEKKLEEENFRVGLTKVFFKAGIVAHLEDLRDQSLA
QLITGLQAQIRWYYQTIERKRRVEKITALKIIQRNIRSWAELRTWVWFKLYGKVKPLV
NSGKIEAQYEKLQETVATLKDTVVQEEEKKRQLQEGAERLNKETADLLAQLEASKGST
REVEERMTAMNEQKVALEGKLADASKKLEVEEARAVEINKQKKLVEAECADLKKNCQD
VDLSLRKVEAEKNAKEHQIRALQDEMRQQDENISKLNKERKNQEEQNKKLTEDLQAAE
EQNLAANKLKAKLMQSLEDSEQTMEREKRNRADMDKNKRKAEGELKIAQETLEELNKS
KSDAENALRRKETELHTLGMKLEDEQAAVAKLQKGIQQDEARVKDLHDQLADEKDARQ
RADRSRADQQAEYDELTEQLEDQARATAAQIELGKKKDAELTKLRRDLEESGLKFGEQ
LTVLKKKGSDAIQELSDQIEQLQKQKGRIEKEKGHMQREFDESCAALDQEAKLRADQE
RIAKGYEVQTSELRLKADEQSRQLQDFVSSKGRLNSENSDLARQVEELEAKIQAANRL
KLQFSNELDHAKRQAEEESRERQNLSNLSKNLARELEQLKESIEDEVAGKNEASRQLS
KASVELDQWRTKFETEGLIGADEFDEVKKRQNQKTSEIQDALDACNAKIVALENARSR
LTAEADANRLEAEHHAQAVSSLEKKQKAFDKVIDEWKKKVDDLYLELDGAQRDARQLS
GEAHKLRGQHDTLADQVEGLRRENKSLSDETRDLTESLSEGGRATHALSKNLRRLEME
KEELQRGLDEAEAALESEESKALRCQIEVSQIRAEIEKRIAEKEEEFENHRKVHQQTI
DSIQATLDSETKAKSELFRVKKKLEADINELEIALDHANKANEDAQKNIRRYLDQIRE
LQQTVDEEQKRREEFREHLLAAERKLAVAKQEQEELIVKLEALERARRVVESSVKEHQ
EHNNELNSQNVALAAAKSQLDNEIALLNSDIAEAHTELSASEDRGRRAASDAAKLAED
LRHEQEQSQQLERFKKQLESAVKDLQERADAAEAAVMKGGAKAIQKAEQRLKAFQSDL
ETESRRAGEASKTLARADRKVREFEFQVAEDKKNYDKLQELVEKLTAKLKLQKKQLEE
AEEQANSHLSKYRTVQLSLETAEERADSAEQCLVRIRSRTRANAEQK"

intron	1790. .1899
	/number=1
exon	1900. .2028
	/number=2
intron	2029. .2083
	/number=2
exon	2084. .2224
	/number=3
intron	2225. .2273
	/number=3
exon	2274. .2748
	/number=4
intron	2749. .2802
	/number=4
exon	2804. .2985
	/number=5
intron	2986. .3033
	/number=5
exon	3034. .3193
	/number=6
intron	3194. .3453
	/number=6
exon	3454. .4745
	/number=7
intron	4746. .4797
	/number=7
exon	4798. .7334
	/number=8
intron	7335. .7381
	/number=8
exon	7381. .7531
	/number=9
intron	7532. .7579
	/number=9
exon	7580. .7870
	/number=10
intron	7871. .8059
	/number=10
exon	8060. .8341
	/number=11
intron	8342. .8387
	/number=11
exon	8388. .>8525
	/number=12

BASE COUNT 3364 a 2316 c 2266 g 2826 t 8 others
ORIGIN

Query Match 3.3%; Score 42; DB 3; Length 10780;
Best Local Similarity 54.0%; Pred. No. 1.9;
Matches 107; Conservative 0; Mismatches 87; Indels 4; Gaps 1;

```
Qy   1045 AAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTTCTCTAGAATTTAATTRTGCTTTTT 1104
          ||  | | || | ||    |    || || |  | || || | | ||| |   | | ||||
Db    777 AAGGATACTACTCAGATATAAAATGTTACGACGTTTTTCAATACTTTTANATTTCATTTT 718

Qy   1105 TTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAAACAACCAGA----AAACCCCTGAA 1160
          |  ||| || | |  | |||| |    || ||||||||  ||| |    ||| |||  ||
Db    717 TGGTTTCTTGTAGCTTATGGGGATGTCCACAATAAAACGTCCACAGCACAAAACCCACAA 658

Qy   1161 GGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACAAACAGCTTTGANCTGAGAGC 1220
          | |  | |     | | |||||  ||| | || |  ||| | | |   | | |
Db    657 TGTACAAGGTATAATTATGCTTATCAAAACTCGAATTGCAACCTGGTCAGGAGTAATTAT 598

Qy   1221 AATTYCAAAAGGCTGCTG 1238
          |||  ||  |  | | |||
Db    597 TATTACAGACAGTTACTG 580
```

RESULT 11
AC006101/c
LOCUS AC006101 155644 bp DNA HTG 04-DEC-1998
DEFINITION *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 10 clone CIT-HSP-1338F24; HTGS phase 2, 3 ordered pieces.
ACCESSION AC006101
NID g3962485
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155644)
AUTHORS Smith,D.R.
TITLE Sequencing of Human Chromosome 10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155644)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA
COMMENT ***
*** WARNING: Phase 2 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 3 contigs for
* which the order is known. The lengths of the gaps have been
* estimated by the submitter but are not known exactly. When
* sequencing is complete, the sequence data presented in this
* record will be replaced by a single finished sequence
* with the same accession number.
* 1 45025: contig of 45025 bp in length
* 45026 48270: gap of unknown length
* 48271 112277: contig of 64007 bp in length
* 112278 115521: gap of unknown length
* 115522 155644: contig of 40123 bp in length.

FEATURES	Location/Qualifiers
source	1. .155644
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="CIT-HSP-1338F24"
	/chromosome="10"

BASE COUNT 43051 a 31296 c 30076 g 44716 t 6505 others
ORIGIN

Query Match 3.3%; Score 41.4; DB 18; Length 155644;
Best Local Similarity 45.0%; Pred. No. 3.2;
Matches 135; Conservative 2; Mismatches 163; Indels 0; Gaps 0;

```
Qy     968 ACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGAAGATGGG 1027
           | |||  | |     ||| ||   | ||  ||  ||     ||  | |   |||   ||
Db    9426 ATTTTTTTATGTTAAAATATTTCAAAATAAAACAAATTTACAACTGTACTTGAATGAGGA 9367

Qy    1028 GAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTTCTCTAGA 1087
           ||   ||    |   ||| |     |||| ||||   |   | | ||  ||   || |
Db    9366 GAAAATGTCCCTGATCAAAAATATGGTCTTAGTCCTAAATGATTGGCTACATTGGTTAAA 9307

Qy    1088 ATTTAATTRTGCTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAAACAACCA 1147
           ||   |||:      |||| | | || |   |||   | |||| || |||   || |  .
Db    9306 ATCACATTGAAAAACTTTTATATGTTGTAATCTTACCGCAAAGGCAGAATTTTACGAGTT 9247

Qy    1148 GAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACAAACAGCT 1207
           |||||||  |  |    || |  ||    |||   || ||||||  |  | |      | |
Db    9246 GAAAACCATTTTAACTGGTCAAGTGAGATAAGTACATTGAAATTAGCCTTAATGTTTGAT 9187

Qy    1208 TTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNTNTCTNAA 1267
           || |         |   : |||    || | || ||   |     ||   | |   | ||
Db    9186 TTTAATGCCACTAACCATTAAATATATGTTCATCTATTTCATTTCTTAGATGTTCATGAA 9127
```

RESULT 12
AC004557/c
LOCUS AC004557 99689 bp DNA PLN 01-DEC-1998
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F17L21, complete sequence.
ACCESSION AC004557
NID g3228389
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 99689)
AUTHORS Shinn,P., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Sun,H., Conway,A., Conway,A., Kurtz,D., Oji,O., Shen,Y.K., Toriumi,M., Vysotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.
TITLE Genomic sequence for Arabidopsis thaliana BAC F17L21
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 99689)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (14-APR-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 99689)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 99689)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT On Jun 16, 1998 this sequence version replaced gi:3097814.
FEATURES Location/Qualifiers

```
     source          1. .99689
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
     CDS             complement(join(71. .672,759. .858,941. .1801))
                     /note="5' end of T7N9.22 (gi|2213602); 3' partial gene
                     that overlaps Sp6 end of T7N9; similar to EST gb|N96069"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.1"
                     /db_xref="PID:g3935158"
                     /translation="MSVHGRFPASPPISLSPSSSSTSPSSQSPSTPPDLKQRVIACLN
                     KLADRDTLALASAELDSIARNLTHDSFSPFLNCIHNTDSSVKSPVRKQCVALLSVLSR
                     YHGDSLTPHLAKMVSTVIRRLRDPDSSVRSACAVATADMSAHVTRQPFASVAKPLIET
                     LIQEGDSNLQIGAALCLAASVDAATDPESEQLRKSLPKIGKLLKSDGFKAKAALLSAV
                     GSIITAGGAGTKPVLDWLVPVLIEFLSSEDWAARKSAAEALGKVATAEDLASQYKKTC
                     TTALESRRFDKVKSVRETMNRALNLWKEVSTDDEASLSPSRSSTDDGNIGCFSSVTRS
                     STIDVGLKSARPKKVTPIMKRSPSLPVNRSYAATRQKENLPKRNQGNMTMLVEEASSV
                     DNKGPHFTPVKKSSEETEEKANSGGPDIIKHTISEKSREDSKVSSFGGLRSGSRVAPC
                     SDDGDSVKNCKDDVEESKKDSEELSLIREQLALIENQQSSLLDLLQVYTWSLISIVSL
                     YVDKLFGYECCR"
CDS join(7691. .7830,7965. .8093,9131. .9329)
/note="hypothetical protein similar to T7N9.23 (gi|2213603)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.2"
/db_xref="PID:g3935159"
/translation="MAGEGSKPKATVDDAYAYIRTVKSTFHNDPDKYDDFMAIMKNFK
ARKIDRNTCIEEVKELLKGHRDLISGFNAFLPKCLEIADWYNIEGRTVESLLNGNTRN
FELHNLIIEIQHWRKKFSNAIIKWSFRSTNKAADRLAKGELENNVTFVSHSTR"
CDS join(10122. .10261,10378. .10624,10700. .10801,10890. .10961)
/note="hypothetical protein similar to T7N9.24 (gi|2213604)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.3"
/db_xref="PID:g3935160"
/translation="MVERRVQVEPTLSDAHSYITAVKEAFHDEPTKYEEFIKLMNDIR
DHGVDKASGIAKLTELIKGHPRLLRGLSFFFPQVNRDIHHEAKRTIILKDKATIPPEA
AYRGAKSTYTKIKQIEPDWENFMNMLKTRFRSLDTHVVESFLKIMIMYDEGKKSEKEV
QEEVVDLLYYHEDLIDKFFRLFNMRK"
CDS join(12453. .12613,12749. .12879,13398. .13479,13874. .13939, 14016. .14285,14382. .14499,14602. .14703,14781. .14852)
/note="hypothetical protein similar to T7N9.25 (gi|2213605)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.4"
/db_xref="PID:g3935161"
/translation="MVGGGSKHVGEGSKPRATKDDAYAYLRAVRAKFHNDSKKYDDFV
AVMTNFKARRIDRDGCIKEVEQLLKGHRDLISGFNAFLPKCLEIKNYYFGAGEDLEVV
YTAKRKLLARSVTKRSSSSSSSSFRDEPAKYRQFLSLLRDRRARRIDKATFFVGLVEL
IKDHLDLLLGFNALLPARFQIPITPAGFQNVVGRSVPPETTIEDATSYLNSVKRAFHD
EPAKYEELLKLLNDIEARRVDAASFIASVEELMKDHQTLLNGFSVFLSAEMKFIRKLK
AKFQGDGSHVADSVLQILRMYSEGNKSKSEAFQEVVPLVQDHEDLVMELIKIIFSDP"
CDS join(15962. .16104,16188. .16481,16578. .16695,16782. .16883, 16966. .17034)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.5"
/db_xref="PID:g3935162"
/translation="MVGKSKSPEFNFEDGMAYFDAVKVALQDTEPEKYQEFVRIFIDY
TANRFGFETLSASLQELLRDHANLCHGSNVLLPFRLQTAIPPEASAEFHINKVVGRSV
PPAVAVPTMDDATSYLNAVKEAFHDEPAKYMEITKLLTDLKARRINAASVIARMEELL
KDHLNLLLGFCVFLSPTRRFITKLKARFLGDGSQVVDSVLQILRMHSEGNKSKDEASQ
EVRALIQGHEDLLMELSEIFSDP"
CDS join(17912. .17988,18081. .18639,18726. .18875)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.6"
/db_xref="PID:g3935163"
/translation="MEEVYHDEPEKFNHILHLIRDYFNHRDDRARITACMGELMRDHL
NLLVRFFDFFPAEASEGLAQLQTIAATSVSPDSTIDDAVSYINTVKEAFHDEPAKYYE
FFQLFYDIRYRYVLLCPSSILFALNFKSQSIQHCEVILFVIYYSLCRLIDVAGGITRV
EELLKAHKNLLVRLNAFLPPEAQRILHLKIEQRAASDINKRKRVASFIGKLKERFQGD
DRHVYESFLEILTMYQEGNKSVNDLYQEVGFLVGTLWLDYTFS"
CDS join(19430. .19507,19610. .19831,20016. .24392)
/note="putative polyprotein (AF039376)"
/codon_start=1
/evidence=not_experimental
```

/product="F17L21.7"
/db_xref="PID:g3935164"
/translation="MFINIVNRVIQVNCWSHTWHEASCLQVFDTKTQTCDPKPIPCSE
TFGGFDYKTLCLDGKFHVGPVPVDKTRIRLADYGGKRAVLWPNNIHSLRKMIWCAESF
KRSLNLNLAATLFFIIFHYFSFSSMASETVITSDSTSLLNVNMTNVTKLTSSNFLMWR
RQVQALLNGYDLTGYIDGSIVVPPATITANGAVTVNPAFKHWQRQDQLIYSALLGAIS
ISVQPILSRTTTSAEIWTKLMDTYAKPSWSHIQQLRQQIKQWKKDTKSIDEFFQGLVM
RFDQLALLGKPMESEEQMEVIVEGLSDDYKQVIDQIQGREVPPSLTEIHEKLLNHEVK
LQAAASSLPISANAASYRPPANNKHNNSNNYRGQNRNNNNRGANSYQQPRNDQPSSRG
YQGKCQICGVFGHSARRCSQLQMSGAYSTPSPSQYPNATVPWQPRANMAAMSYNPWLL
DSGATHHLTTDLNNLALHQPYNGGEEVTIADGSTLPITHTGSSTLSTQSRSLALNNIL
YVPNLHKNLISVYKLCNANKVSVEFFPAHFQVKDLSTGARLLQGRTKDELYEWPVPSN
TPISLFASPTPKTTLPSWHSRLGHPSPPVLKSLVSQFSLPVSNSSQKHFPCSHCLINK
SHKLPFYSNTIISYTPLEYVYSDVWTSPVTSVDNFKYYLILVDHYTRYTWLYPLKQKS
QVRETFVAFKALVENRFQTKIRTLYSDNGGEFIALRQFLLTHGISHLTSLPHTPEHNG
IAERKHRHILETGLTLLTQASIPTSYWTYAFGTAVYLINRLPSSVLNNESPYSKLFKT
SPNYLKLRVFGCSCFPWLRPYTNHKLERRSQPCVFLGYSLTQSAYLCLDRSSGRVYTS
RHVQFVEDQFPFSISDTHSVSNSSPEEASPSCHQPPSRIPIQSSSPPLVQAPSSLPPL
SSDSHRRPNAETSSSSSSTNNDVVVSKDNTQVDNRNNFIGPTSSSSAQSQNNSNPSSS
IQTQNEPNPSPSPTPQNSSPESSPSSSTSATSTVPNPPPPPPTNNHPMRTRAKNHITK
PKTKLSLLAKTVQTRPQIPNTVNQALRDEKWRNAMGEEINAQIRNNTFELVPPKPNQN
VISTKWIFTLKYLPNGTLDRYKARLVARGFRQQYGLHYSETFSPVVKSLTIRLVLQLA
VSRSWTIKQLDVNNAFLQGTLTDEVYVTQPPGFIDPDRPHHVCRLKKALYGLKQAPRA
WYQELRNFVCSLGFTNSLADTSVFVYINDIQIVYCLVYVDDIIVTGSSDALVMAFITA
LSRRFSLKDPTDLVYFLGIEATRTSQGLHLMQHKYVYDLLSRMKMLDAKPVSTPMATH
PKLSLYSGIALDEPGEYRTVIGSLQYLAFTRPDIAYAVNRLSQFMHRPTDIHWQAAKR
VLRYLAGTATHGILLRSNSPLSLHAFSDADWAGDNDDFVSTNAYIVYLGSTPIAWSSK
KQKGVARSSTEAEYRAVANTTSEIRWVCSLLTELGITLPKMPVIYCDNVGATYLSANP
VFHSRMKHLALDYHFIRDNVSAGALRVSHISTHDQLADALTKPLPRQHFLQFSSKIGV
SKLPPS"
CDS join(25673. .26084,26445. .26461)
/note="unknown; similar to ESTs gb|T42386, gb|H36247,
gb|N38423, gb|T21830, and gb|AA585725"
/codon_start=1
/evidence=not_experimental
/product="F17L21.8"
/db_xref="PID:g3935165"
/translation="MMRRQRDEQQSRVFYDLSSLVLNLLRSPPMPISLPDHFPDPPAR
IRPPFSSHRFSSSSPSLAHISPSGFASLLLGISVALMLCGSVTFFIGFLLMPWVLALI
LVFYVAGIVSAISMVGRSILCYVLTPPSPSGKEISEWKLL"
CDS join(27209. .27349,27618. .28079)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.9"
/db_xref="PID:g3935166"
/translation="MESSEDVENLSRAIEKLLHEKRKREASGDAFIEDADDQLFLSSL
ISQLESPNPKGKEDVITGEEEEKEEESADSSPSKGKSEGQRQLEESIEEIAKDIKKVK
KQNKITHVLLSATIILTLVWQLSEYSMIFMLKDRISHPVRSIGGMLNGMFKGKLRPIK
NQLAGTSNSNDQNNHGNGSHTGPQLQVPELLREFGFDDEE"
CDS complement(join(28500. .28571,29379. .29675))
/note="similar to nuclear transport factor 2; similar to
ESTs gb|T21347, gb|N96312, and emb|F20149"
/codon_start=1
/evidence=not_experimental
/product="F17L21.10"
/db_xref="PID:g3935167"
/translation="MDPDAVAKAFVEHYYSTFDANRPGLVSLYQEGSMLTFEGQKIQG
SQNIVAKLTGLPFQQCKHNITTVDCQPSGPAGGMLVFVSGNLQLAGEQHALKFSQMFH
LISNQGNYYVFNDIFRLNYA"
CDS join(31665. .32171,32453. .32734,32937. .33106,33195. .33690,
33784. .33835,33950. .34179,34268. .34366,34453. .35558,
35648. .35827,35899. .35986,36081. .36149)
/note="similar to sensory protein barA (sp|P26607);
similar to histidine kinase A (gi|1136289); similar to
ESTs gb|AA712891 and gb|AA042438"
/codon_start=1
/product="F17L21.11"
/db_xref="PID:g3935168"
/translation="MSLFHVLGFGVKIGHLFWMLCCWFVSWFVDNGIEDKSGLLVGSV
GDLEKTKMTTLKKKNKMWFWNKISSSGLKIPSFSYQFLGSVKFNKAWWRKLVVVWVVF
WVLVSIWTFWYFSSQAMEKRKETLASMCDERARMLQDQFNVSMNHVQAMSILISTFHH
GKIPSAIDQRTFSEYTDRTSFERPLTSGVAYAMRVLHSEREEFERQQGWTIRKMYSLE
QNPVHKDDYDLEALEPSPVQEEYAPVIFAQDTVSHVVSLDMLSGKEDRENVLRARSSG
KGVLTAPFPLIKTNRLGVILTFAVYKRDLPSNATPKERIEATNGYLGGVFDIESLVEN
LLQQLASKQTILVNVYDITNHSQPISMYGTNVSADGLERVSPLIFGDPLRKHEMRCRY
LQLAHTYVCNFFLFARIQVLTFCCELLPLCRFKQKPPWPVLSMVTSFGILVIALLVAH
IIHATVSRIHKVEEDCDKMKQLKKKAEAADVAKSQFLATVSHEIRTPMNGVLGMLHML
MDTELDVTQQDYVRTAQASGKALVSLINEVLDQAKIESGKLELEEVRFDLRGILDDVL
SLFSSKSQQKGVELAVYISDRVPDMLIGDPGRFRQILTNLMGNSIKFTEKGHIFVTVH
LVDELFESIDGETASSPESTLSGLPVADRQRSWENFKAFSSNGHRSFEPSPPPDINLIV
SVEDTGVGIPVEAQSRIFTPFMQVGPSISRTHGGTGIGLSISKCLVGLMKGEIGFSST
PKVGSTFTFTAVFSNGMQPAERKNDNNQPIFSEFRGMKAVVVDHRPARAKVSWYHFQR
LGIRVEVVPRVEQALHYLKIGTTTVNMILIEQEIWNREADDFIKKLQKDPLFLSPKLI
LLANSVESSISEALCTGIDPPIVIVKPLRASMLAATLQRGLGIGIREPPQHKGPPALI
LRNLLLGRKILIVDDNNVNLRVAAGALKKYGADVVCAESGIKAISLLKPPHEFDACFM
DIQMPEMDGFEATRRIRDMEEEMNKRIKNGEALIVENGNKTSWHLPVLAMTADVIQAT
HEECLKCGMDGYRIIRGDGAGRRTAKETQKLVSPMHLEVVEKQIDDCKKSVYRKAGVT
RPRT"
CDS join(36949. .36951,37408. .37564,37737. .37783)
/note="unknown; similar to EST gb|AA650671 and gb|T20610"
/codon_start=1
/evidence=not_experimental
/product="F17L21.12"
/db_xref="PID:g3935169"
/translation="MTTSKRLADRKIEKFDKNILKRGFVPETTTKKGKDYPVGPILLG
FFVFVVIGSSLFQIIRTATSGGMA"
CDS join(39626. .39777,40491. .41742)
/note="similar to fim protein; similar to ESTs gb|T42445,
gb|T76780, gb|AA650733, and emb|Z17748"
/codon_start=1
/evidence=not_experimental
/product="F17L21.13"
/db_xref="PID:g3935170"
/translation="MEEELAMLRQLIGQLQELLHNGSPPPPSSSSSLSSSSPSFLVLH
HPQYQNGWCLPCIEDTSADDCCDIVMAGGKRPGIFKMLETVKPPVKRTRKERTQGKSC
TEVDEISGNMDQEIWQEFPQDLFEDVVSRLPMATFFQFRAVCRKWNALIDSDSFSRCF
TELPQTIPWFYTITHENVNSGQVYDPSLKKWHHPIIPALPKKSIVLPMASAGGLVCFL
DIGHRNFYVSNPLTKSFRELPARSFKVWSRVAVGMTLNGNSTSHGYKVLWVGCEGEYE
VYDSLSNVWTKRGTIPSNIKLPVLLNFKSQPVAIHSTLYFMLTDPEGILSYDMVSGKW
KQFIIPGPPDLSDHTLAACGERLMLVGLLTKNAATCVCIWELQKMTLLWKEVDRMPNI
WCLEFYGKHIRMNCLGNKGCLILLSLRSRQMNRLITYNAVTREWTKVPGCTVPRGRKR
LWIACGTAFHPSPTARA"
CDS complement(join(42083. .42239,42509. .42669,43186. .43188))
/note="unknown; similar to ESTs gb|T20610 and
gb|AA586199"
/codon_start=1
/evidence=not_experimental
/product="F17L21.14"
/db_xref="PID:g3935171"
/translation="MTTSKRLADRKIEKFDKNILKRGFVPETTTKKGKDYPVGPILLG
FFVFVVIGSCESLGLQLAEAWHKPKHSQARRSSFQNVESSSVSFLFLSLYCRMCQLSV
YYIM"
CDS join(46024. .46685,46846. .46967,47100. .47351,47455. .47600)
/note="putative squamosa-promoter binding protein
(AC002561)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.15"
/db_xref="PID:g3935172"
/translation="MDCNMVSSSQWDWEHLIMSNPSRTEDDSKQLPTEWEIEKGEGIE
SIVPHFSGLERVSSGSATSFWHTAVSKSSQSTSINSSSPEAKRCKLASESSPGDSCSN
IDFVQVKAPTALEVSVASAESDLCLKLGKRTYSEEYWGRNNNEISAVSMKLLTPSVVA
GKSKLCGQSMPVPRCQIDGCELDLSSAKGYHRKHKVCEKHSKCPKVSVSGLERRFCQQ
CSRFHAVSEFDEKKRSCRKRLSHHNARRRKPQGVFSMNPERVYDRRQHTNMLWNGVSL
NARSEEMYEWGNNTYDTKPRQTEKSFTLSFQRGNGSEDQLVASSSRMFSTSQTSGGFP
AGKSKFQLHGEDVGEYSGVLHESQDIHRALSLLSTSSDPLAQPHVQPFSLLCSYDVVP
K"
CDS complement(join(49339. .49493,49603. .49851,49967. .50088,
50218. .50882))
/note="similar to squamosa-promoter binding protein
(gi|2673911)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.16"
/db_xref="PID:g3935173"

/translation="MDCNMVSSFPWDWENLIMSNQSKTENEKKQQSTEWEFEKGEGIE
SIVPDFLGFEKVSSGSATSFWHTAVSKSSQSTSINSSSPEDKRCNLASQSSPGDSSSN
IDFLQVKPSTALEVPIASAESDLCLKLGKRTYSEEFWGRNNNDLSAVSMNLLTPSVVA
RKKTKSCGQSMQVPRCQIDGCELDLSSSKDYHRKHRVCETHSKCPKVVVSGLERRFCQ
QCSRFHAVSEFDEKKRSCRKRLSHHNARRRKPQGVFPLNSERVFDRRQHTSMLWNGLS
LNTRSEEKYTWGTTYETKPTQMESGFTLSFQRGNGSEDQLFTGSTLSFSAFQTSGGFS
AGKSNIQLPDKGVGECSGGLHESHDFYSALSLLSTTSDSQGIKHTPVAEPPPIFGTFP
SHFI"

CDS complement(join(52788. .52954,53423. .53556,53832. .53974))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.17"
/db_xref="PID:g3935174"
/translation="MKDRMERLLILPFSLGCSTQSCVAVATTHQQKKPNQLIKSTFLL
TFLLVESDAKGSFRKDYTKMENNNNGANISDGIYRIIRSFKSFSHFFIRYEEETKERE
AEMEIGFPTDVKHLSHIGVDGTMTTFDNTSSSFPFSGFHLTGTVV"

CDS complement(join(55014. .55070,55246. .55329,55703. .55816,
56001. .56104,56456. .56705))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.18"
/db_xref="PID:g3935175"
/translation="MACKTILRSVFVSESRRTSGARRCFFLPPSPASVPVHGLFPAPK
SLCFNGFASVPERATRLNCSHNDQSDQGPPQEAVLKAISEVSKTDGRVGKTTNMIIGG
TVADDSAKDWLELDQKVNTYPTERGFTAIGTGGNDFVHAMVVAVESVIDRQIPEDCVK
QTLSSKGKYVSVNIGPIQVVSSEQVQAVYNAMRRDERMKYFL"

CDS complement(join(57354. .57485,57663. .57761,57936. .57983,
58075. .58244,58331. .58409,58699. .58797))
/note="similar to TOM20 (X92491); similar to ESTs
gb|T444475, emb|Z26777, and emb|Z26778"
/codon_start=1
/product="F17L21.19"
/db_xref="PID:g3935176"
/translation="MEFSTADFERLIMFEHARKNSEAQYKNDPLDSENLLKWGGALLE
LSQFQPIPEAKLMLNDAISKLEEALTINPGKHQALWCIANAYTAHAFYVHDPEEAKEH
FDKATEYFQRAENEDPGNDTYRKSLDSSLKAPELHMQFMNQGMGQQILGGGGGGGGGG
MASSNQSSKKKKRNTEFTYDVCGWIILACGIVAWVGMAKSLGPPPPAR"

CDS join(59115. .59117,59214. .59250,59369. .59409,59889. .60119,
60204. .60464,60563. .60610)
/note="similar to 60S ribosomal protein L17-2 (sp|P35267);
similar to ESTs gb|L33542 and gb|AA660016"
/codon_start=1
/evidence=not_experimental
/product="F17L21.20"
/db_xref="PID:g3935177"
/translation="MVKYSQEPDNITKSCKARGADLRVHFKNTRETAHAIRKLPLNKA
KRYLEDVIAHKQAIPFTRFCRGVGRTAQAKNRHSNGQGRWPAKSAQFVLDLLKNAESN
AEVKGLDVDALFISHIQVNQAAKQRRRTYRAHGRINRMSSFLFDCFYNVCFFIHIVLT
LDLILSYAAYMSNPCHIELILSEKEEPVKKEPETQLAAKSKKGASS"

CDS join(61111. .61402,61486. .61688,61772. .61879,61963. .62628)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.21"
/db_xref="PID:g3935178"
/translation="MESGLISVDRWRNGSQAYFLTHIHSDHTRGLSGGWSQGPLYCSR
TTASLFPSRFPGFDLSLLRVVPLFSWTSLSLRSPSSGSTVRLHLMAIDAHHCPGSIMF
LFRGDFGCFLYTGDFRWDSDASDEARTTLVAAIDEFPVDILYLDNTYCNPIYSFPSRL
VAVQLVADIIASHPSHDIIIAVDSLGKEDLLVHVSRILNIKIWVWPERLRTMHLLGFQ
DIFTTDTSLTRVRAVPRYSFSIQTLEGLNTMCPTIGIMPSGLPWVKRPFKGDDKLSGS
FLTASMKNETVSAKKELEAAAVHKFHDYMYSVHYSDHSCYEEIGEFIKLVKPKSMKGI
VVSSSSYVDPLYYFGRICGANQPPQVLLMRPDIAEEFRAVRIKSYSASDKTRMLAKEK
RWKRDSHSSLKRNKKRARIQVKCAKILEVD"

CDS join(62993. .63429,63730. .63953,64026. .64405)
/note="similar to MIPP proteins (gi|2191178, sp|P28575)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.22"
/db_xref="PID:g3935179"
/translation="MVVESSSSPIIPGLTDDVAELCVSKIPRSSFQITSQVCRRWRSF
LRSQHFAAVRKLTGTVEEFLCVLMESECGRDVYWEVFDASGNKLGQIPPVPGPLKRGF
GVAVLDGGKIVFFGGYTEVEGSGINSTTVSASADVYEFDPANNSWRKLAGMNIPRYNF
AFAEVNGLLYVIRGYSTDTYSLSNAEVYNPKTNQWSLMHCPNRPVWRGFAFAFSSKLY
AVGNGSRFIDIYDPKTQTWEELNSEQSVSVYSYTVVRNKVYFMDRNMPGRLGVFDPEE
NSWSSVFVPPREGGFWVRLGVWNNKVLLFSRVCGHETLMYDLDKEKGSKWRVCDQIKP
SASQLASVLINF"

CDS complement(join(64930. .65189,65274. .65411,65541. .67868,
68016. .68726,68873. .68971,69070. .69160,69275. .70105,
70275. .70345,70747. .70813))
/note="unknown; similar to ESTs gb|N96209, gb|R87018,
gb|H76168, and gb|AA395332"
/codon_start=1
/evidence=not_experimental
/product="F17L21.23"
/db_xref="PID:g3935180"
/translation="MAEGKFDLPDDLIFSKSSDQLKELASDNSIPLSPQWLYTKSSEY
KMDVRSPTPVPMGNPSDPNPKDAWRLDAPEDKKDWKKIVHENETSRRWREEERETGLL
GARKVDRRKTERRIDSVSSRETGDIKNAAASDRWNDVNSRAAVHEPRRDNKWSSRWGP
DDKEKEARCEKVDINKDKEEPQSESQSVVSNVRATSERDSDTRDKWRPRHRMESQSGG
PSSYRAAPGFGLDRGRAEGPNLGFTVGRGRASTIGRGSSTSLIGAGSALSPVFRYPRG
KLLDMYRKQKPDSSLGRILTEMDEVASITQVALIEPLAFIAPDAEEEANLNGIWKGRI
ISSEVYTSSGEESLGGNSLLKCRIPESGETKVDGALLGFMNGDNGSMKNNDSGLLGSH
NGGLGAASSVPRLNSVASESYGSGGAGYQLSHGSPEAVRSVFTKSSVLDGSESVVGSF
EQAYTGKLQQPDTEVDHSEGAMPPEEFLFLYIDPQGVIQGPFIGSDIISWFEQGFFGT
DLQVRLASAPEGTPFQDLGRVMSYIKAESVHAHISDQKSELEETSLKANSEAGGSVAH
VAESNDSSSLTGISRSFSVYNNPSGQDNFQRKSESEVYGRPPHAEDQSFLDFSAQDEE
IVFPGRARVSGYASSVKSSTSMHDALMEFSGHSDIPVEVTTAATRNQNENKLHPFGVL
WSELEGGSTPVNPLPNRSSGAMGEPSCSIENRPINSRRNSQIDPNISLDALSGNRMSQ
FEHESNFFNHGDQLPSNQHHQQHFQNRDMLSHLHIGDQDLEHLITLQLQQQQKIQMQQ
QQKIQLQQQQKIQLQQHQLEQEHQLHQKLLQEQQQSHARQLHFQQILQGQTPDTRFGQ
SHDFPRSNSVDQMLLEQQMLNELQKSSGHPSQNFAPYIEQHAAGNFGRFTHEGHQREL
LEQLFSTQMQSQYGQKQSQYGQMQSQHGQLQSEPIRSLEYQLLQQEQLMQLANGVRHN
TLLEEQRHIDPLWPSDHSDQLLRTHPGIHRSHSSAGFRPLDFHQQQQRPHFEDQFSQL
ERNRSYQQQLRLELLEHGLPFERSASGLNLDAVNGLGLSQGLELRDATAHMQSSGRLG
NSTPGFSHQNPRIPLGESHFSHLEPTEGRWSGADTQLAGDWAESQFRRSNMDTEHDKM
RSEIRRLGEDPNSWMVGGSTDDKSKQLFMELLHQRPGHQSAESPNMNRGYPYDRMVPS
GLTPGIQTLGGLSDHGGNQNVSSAFGDRSFSDEQVNRVPGYGNNMGSLHHNSSLLSGI
IDAGRSTQNETQAFSNMFGMNKDANDINTWNNVPPKNEGMGRMMSYDAQDRMGKQAVL
DSLIQEELPVGTPGQQSSFNISGMLSFRVIASSKIFCYFSNFMVWKVDFYFVIMKAYP
DILLLQTDTVTTWLEKIEGRIVPSHGQNSVLLKRPPSSHSSSSHEGLLERMSDTASRA
AASSYSGIEGGVRRESGAAGNKGSTSEAASFSEMLKKSNSMKKVAAESTDATEGSKGG
GGKKKGKKGRQIDPALLGFKVTSNRILMGEIHRADDF"

CDS complement(join(73150. .73663,73910. .74151,74405. .74674))
/note="similar to multiple exostoses type II protein
EXT2.I (U72263); similar to ESTs dbj|D39982, gb|L37635,
and dbj|C28418"
/codon_start=1
/evidence=not_experimental
/product="F17L21.24"
/db_xref="PID:g3935181"
/translation="MFAAEIFMHRFLLSSPVRTRNPDEADWFYTPIYPTCDLTPTGLP
LPFKSPRMMRSSIQLISSNWPYWNRTEGADHFFVVPHDFGACFHYQEEKAIERGILPL
LQRATLVQTFGQRNHVCLDEGSITIPPFAPPQKMQAHFIPPDIPRSIFVYFRGLFYDV
NNDPEGGYYARGARAAVWENFKNNPLFDISTDHPTTYYEDMQRAIFCLCPLGWAPWSP
RLVEAVVFGCIPVIIADDIVLPFADAIPWEEIGVFVAEKDVPELDTILTSIPTEVILR
KQRLLANPSMKRAMLFPQPAQPGDAFHQILNGLARKLPHDKSIYLKTGEKALNWTAGP
VADLKPW"

CDS join(76306. .76387,76478. .76588,76756. .76848,77250. .77399,
77484. .77542,77636. .77692)
/note="putative adenine phosphoribosyltransferase 1
(sp|P31166); similar to ESTs gb|N65531, gb|R90631,
gb|T21275, and gb|AA713070"
/codon_start=1
/evidence=not_experimental
/product="F17L21.25"
/db_xref="PID:g3935182"
/translation="MATEDVQDPRIAKIASSIRVIPDFPKPGIMFQDITTLLLDTEAF
KDTIALFVDRYKDKGISVVAGVEARGFIFGPPIALAIGAKFVPMRKPKKLPGKVISEE
YSLEYGTDTIEMHVGAVEPGERAIIIDDLIATGGTLAAAIRLLERVGVKIVECACVIE
LPELKGKEKLGETSLFVLVKSAA"

CDS join(78862. .79059,79153. .79411,79507. .80678,80747. .80891,
80999. .81309)
/note="hypothetical protein"
/codon_start=1

/evidence=not_experimental
/product="F17L21.26"
/db_xref="PID:g3935183"
/translation="MLCACSGEQFRFEDQPGSPESLATRDFSASGLSSRNGGGDWDSK
LEDIQVDEAESTLKEALSLNYEEARALLGRLEYQRGNFDAALQVFKGIDIKVLTPRII
KAIVEKTLPCKPRSKAVIVPPTTMSMHSVSLLLEAILLKARSLEELGSYKEAAEECKI
ILDVVENALPSGMPDGISGFAKLQDIFQKALELLPLLWKKAGNHHETIASYRRALSRP
WNLDPQRLAVTQKSLALVLLYGSVEACPKDNIEEAIVLLMLLVKKMVVGDIQWDPELM
DHLTYALSMTGQFEVLANYLEQTLPGVYTRGERWYLLSLCYSAAGIDKAAINLLKMAL
GPSESRQIPHIPLLLFGAKLCSKDPKHSRDGINFAHRLLDLGNSQSEHLLSQAHKFLG
VCYGNAARSSKLDSERVFLQKKSLFSLNEAAKRGKADPELDVIFNLSVENAVQRNVQA
ALDGAVEYSSMVGGVSTKGWKHLAIVLSAEKRLKDAESILDFTMEEAGDIEKIELLRL
KAVLQMAQEQPKKAMKTCSSLLGLIRAQEKSEQSESLLQKFETEAWQDLASVYGKLGS
WSDAETCLEKARSMCYYSPRGWNETGLCLEAKSLHEEALISFFLSLSIEPDHVPSIVS
IAEVMMKSGDESLPTAKSFLMNALRLDPRNHDAWMKLGHVAKKQGLSQQAAEFYQAAY
ELELSAPVQSFI"
CDS complement(join(81833. .82924,84388. .84580,84662. .84922,
85075. .85256,85340. .85917,86264. .86411,86652. .86814,
86888. .87001,87154. .87217,87303. .87594,87699. .87842,
87945. .88103))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.27"
/db_xref="PID:g3935184"
/translation="MFEYRCSSIDWKPSPVVALATSADGSQVAAAREDGSLELWLVSP
GASSWHCQLIIHGDPKSRISHLAWFGTGSKGSSSARLFSSSIDGSISEWDLFDLKQKV
VLDSIGFSIWQIALAPISIEAESKEAKLIQNGYLSEKSDDEEESGVEDDSELDELDEK
VEALDRHLAAACDDGCVRIYYISESDKLTYYRSLPRVTGRALSVTWSADAQRIFSGSS
DGLIRCWDANLCHEVYRITVGLGGLGNGSELCIWSLLSLRYGVLVSGDSTGSVQFWDS
QHGTLLQAHSNHKGDVNALAASPSHNRVFSAGADGQVILYKLSGGTSRSQDFKPSSAQ
KWDYIGCVRAHTHDIRALTVAVPISWEGSIPDRNAKVTSPKQRRKEPAGFSYHKWAHL
GVPMLISAGDDAKLYAYSVQEFTKFPPHDICPAPQRVPMQMVHNTVFNQTSLLLVQDS
CSLDILRIHISNDSSGRVSTKPLVRVKSKDARKIICSAISNTGSLFAYSDQVRPSLFE
LKKNKIGKNPWSANRKRLPNLPFAHSMVFSCDCSRLIIAGHDRRIYIVDVGSMELLHS
FTPRQEGQEGESPPHEPPITKLYTSSNDDWLAAINCFGDIYVFNLETQRQHWFISRLD
GASVAAAGFHPRDNNVIVISTSSNQVFALDVEARELSKWSLLQTFCLPKSYQNFPGEV
VGLSFSPSPCSSSVIIYSSRAKCLIEFGKPAEQDEDTDTPCNLSEKLEGKLASISMKL
GNGAQKRRLEEYQKESKSNKRKKFEMFFFFIILDVMIKILNPHSHHSQTTTLKTAEIL
SKYRPIAPKPGTPRVNDDDPSSSMSHKISQSPYLRNLWPQLQARPTRTRKRGRGGMGP
SSLAMKRPKSSCGSSSTSTISTQRVLGPIKTLSFQAFTHHRLPNLPQVGYGFENGVSS
TLVTLPLLQCSPPSSKCMEPEIKGKGVIDLNKTAEVIQERDFLTQLQGPITTTTTATT
SRVISPQPIRPVCSKINVAYINPLTNPSPTSQTSKKSPREVEEDVESDDLPSVITDSN
SRVRLVNSAYKEMMGQPECSWLDSMVRGKRICGEVMINFCESKIPVMTENNGFSCWVR
IDWGRDGKEEYMHAFCDVTKLACDSKDYVFTWRFHTTTDRR"
CDS join(88492. .88892,89156. .90053)
/note="similar to lecithin:cholesterol acyltransferase
precursor (M26268)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.28"
/db_xref="PID:g3935185"
/translation="MKKISSHYSVVIAILVVVTMTSMCQAVGSNVYPLILVPGNGGNQ
LEVRLDREYKPSSVWCSSWLYPIHKKSGGWFRLWFDAAVLLSPFTRCFSDRMMLYYDP
DLDDYQNAPGVQTRVPHFGSTKSLLYLDPRLRDATSYMEHLVKALEKKCGYVNDQTIL
GAPYDFRYGLAASGHPSRVASQFLQDLKQLVEKTSSENEGKPVILLSHSLGGLFVLHF
LNRTTPSWRRKYIKHFVALAAPWGGTISQMKTFASGNTLGVPLVNPLLVRRHQRTSES
NQWLLPSTKVFHDRTKPLVVTPQVNYTAYEMDRFFADIGFSQGVVPYKTRVLPLTEEL
MTPGVPVTCIYGRGVDTPEVLMYGKGGFDKQPEIKYGDGDGTVNLASLAALKVDSLNT
VEIDGVSHTSILKDEIALKEIMKQISIINYELANVNAVNE"
CDS complement(join(90258. .90471,90595. .90775,90859. .90958))
/note="similar to MYB transcription factor isolog
(AC002335)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.29"
/db_xref="PID:g3935186"
/translation="MEWSLPVDLQEEILSRVPAKSLARWKSTPKQWKGPISIEFLHLL
RSTLKFPLLNLHLKSKSIYVTSFTAMVFCYAPPSTKDSRFGIHVQGKPSGFLADEQNK
VVVCCKWRSDSINTIYFVGENKHVQVDQRRGDLTLGQSCSFLMNYVPSFVQIQQGTLL
APRT"
CDS complement(join(95514. .95715,95789. .97565,98312. .98357,
99010. .99108))
/note="similar to kinesin light chain (Z25664); similar
to EST gb|R90451"
/codon_start=1
/evidence=not_experimental
/product="F17L21.30"
/db_xref="PID:g3935187"
/translation="MQLASYRGQGKIADFGFKKPRWVEGKLLKLNGKCRPYITIGIEN
RNRPYSCLFEIKKMEGGSVNESHSNADQMFDTTIEELCKNLCELQSSNQSPSRQSFGS
YGDESKIDSDLQHLALGEMRDIDILEDEGDEDEVAKPEEFDVKSNSSNLDLEVMPRDM
EKQTGKKNVTKSNVGVGGMRKKKVGGTKLQNGNEEPSSENVELARFLLNQARNLVSSG
DSTHKALELTHRAAKLFEASAENGKPCLEWIMCLHVTAAVHCKLKEYNEAIPVLQRSV
EIPVVEEGEEHALAKFAGLMQLGDTYAMVGQLESSISCYTEGLNIQKKVLGENDPRVG
ETCRYLAEALVQALRFDEAQQVCETALSIHRESGLPGSIAEAADRRLMGLICETKGDH
ENALEHLVLASMAMAANGQESEVAFVDTSIGDSYLSLSRFDEAICAYQKSLTALKTAK
GENHPAVGSVYIRLADLYNRTGKVREAKSYCENALRIYESHNLEISPEEIASGLTDIS
VICESMNEVEQAITLLQKALKIYADSPGQKIMIAGIEAQMGVLYYMMGKYMESYNTFK
SAISKLRATGKKQSTFFGIALNQMGLACIQLDAIEEAVELFEEAKCILEQECGPYHPE
TLGLYSNLAGAYDAIGRLDDAIKLLGHVVGVREEKLGTANPVTEDEKRRLAQLLKEAG
NVTGRKAKSLKTLIDSDLTSSSALR"
BASE COUNT 31054 a 18594 c 18732 g 31309 t
ORIGIN

Query Match 3.2%; Score 41; DB 9; Length 99689;
Best Local Similarity 49.5%; Pred. No. 3.9;
Matches 136; Conservative 0; Mismatches 135; Indels 4; Gaps 1;

Qy 662 ACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAAG 721
Db 12114 ACAAACCAGAAAATTTACTATGGGTATAAGGTTCACCACTAGAACTTGAATGATTTTAAT 12055

Qy 722 ACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAAT----CCA 777
Db 12054 TTGAGCTGTATAGGTCTTTATATATATATAAAAAAAAATAGACCTTTTGCAATTGTATCA 11995

Qy 778 TAACATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATT 837
Db 11994 TAATACTAAATATAATAAACACATGTTAGTTGTAAATAAAAATTTCACATTAATTACAAT 11935

Qy 838 TGTTAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTAC 897
Db 11934 TTACAGATTTTTATGTAAATTTTAATTATACATATATTTATGTAAATTTAGTAATAAATA 11875

Qy 898 ATGGAGACAGGTCAACCAAAGAGGGAGCTAGGCAA 932
Db 11874 ATTTAAATTGTAAAGCTTTACAAGCATCTGTGAAA 11840

RESULT 13
AC004557/c
LOCUS AC004557 99689 bp DNA PLN 01-DEC-1998
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F17L21, complete sequence.
ACCESSION AC004557
NID g3228389
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 99689)
AUTHORS Shinn,P., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Sun,H., Conway,A., Conway,A., Kurtz,D., Oji,O., Shen,Y.K., Toriumi,M., Vysotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.
TITLE Genomic sequence for Arabidopsis thaliana BAC F17L21
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 99689)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (14-APR-1998) Arabidopsis thaliana Genome Center,

```
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   3  (bases 1 to 99689)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   4  (bases 1 to 99689)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT     On Jun 16, 1998 this sequence version replaced gi:3097814.
FEATURES             Location/Qualifiers
     source          1. .99689
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
     CDS             complement(join(71. .672,759. .858,941. .1801))
                     /note="5' end of T7N9.22 (gi|2213602);  3' partial gene
                     that overlaps Sp6 end of T7N9;  similar to EST gb|N96069"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.1"
                     /db_xref="PID:g3935158"
                     /translation="MSVHGRFPASPPISLSPSSSSTSPSSQSPSTPPDLKQRVIACLN
                     KLADRDTLALASAELDSIARNLTHDSFSPFLNCIHNTDSSVKSPVRKQCVALLSVLSR
                     YHGDSLTPHLAKMVSTVIRRLRDPDSSVRSACAVATADMSAHVTRQPFASVAKPLIET
                     LIQEGDSNLQIGAALCLAASVDAATDPESEQLRKSLPKIGKLLKSDGFKAKAALLSAV
                     GSIITAGGAGTKPVLDWLVPVLIEFLSSEDWAARKSAAEALGKVATAEDLASQYKKTC
                     TTALESRRFDKVKSVRETMNRALNLWKEVSTDDEASLSPSRSSTDDGNIGCFSSVTRS
                     STIDVGLKSARPKKVTPIMKRSPSLPVNRSYAATRQKENLPKRNQGNMTMLVEEASSV
                     DNKGPHFTPVKKSSEETEEKANSGGPDIIKHTISEKSREDSKVSSFGGLRSGSRVAPC
                     SDDGDSVKNCKDDVEESKKDSEELSLIREQLALIENQQSSLLDLLQVYTWSLISIVSL
                     YVDKLFGYECCR"
     CDS             join(7691. .7830,7965. .8093,9131. .9329)
                     /note="hypothetical protein similar to T7N9.23
                     (gi|2213603)"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.2"
                     /db_xref="PID:g3935159"
                     /translation="MAGEGSKPKATVDDAYAYIRTVKSTFHNDPDKYDDFMAIMKNFK
                     ARKIDRNTCIEEVKELLKGHRDLISGFNAFLPKCLEIADWYNIEGRTVESLLNGNTRN
                     FELHNLIIEIQHWRKKFSNAIIKWSFRSTNKAADRLAKGELENNVTFVSHSTR"
     CDS             join(10122. .10261,10378. .10624,10700. .10801,10890. .10961)
                     /note="hypothetical protein similar to T7N9.24
                     (gi|2213604)"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.3"
                     /db_xref="PID:g3935160"
                     /translation="MVERRVQVEPTLSDAHSYITAVKEAFHDEPTKYEEFIKLMNDIR
                     DHGVDKASGIAKLTELIKGHPRLLRGLSFFFPQVNRDIHHEAKRTIILKDKATIPPEA
                     AYRGAKSTYTKIKQIEPDWENFMNMLKTRFRSLDTHVVESFLKIMIMYDEGKKSEKEV
                     QEEVVDLLYYHEDLIDKFFRLFNMRK"
     CDS             join(12453. .12613,12749. .12879,13398. .13479,13874. .13939,
                     14016. .14285,14382. .14499,14602. .14703,14781. .14852)
                     /note="hypothetical protein similar to T7N9.25
                     (gi|2213605)"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.4"
                     /db_xref="PID:g3935161"
                     /translation="MVGGGSKHVGEGSKPRATKDDAYAYLRAVRAKFHNDSKKYDDFV
                     AVMTNFKARRIDRDGCIKEVEQLLKGHRDLISGFNAFLPKCLEIKNYYFGAGEDLEVV
                     YTAKRKLLARSVTKRSSSSSSSSFRDEPAKYRQFLSLLRDRRARRIDKATFFVGLVEL
                     IKDHLDLLLGFNALLPARFQIPITPAGFQNVVGRSVPPETTIEDATSYLNSVKRAFHD
                     EPAKYEELLKLLNDIEARRVDAASFIASVEELMKDHQTLLNGFSVFLSAEMKFIRKLK
                     AKFQGDGSHVADSVLQILRMYSEGNKSKSEAFQEVVPLVQDHEDLVMELIKIIFSDP"
     CDS             join(15962. .16104,16188. .16481,16578. .16695,16782. .16883,
                     16966. .17034)
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.5"
                     /db_xref="PID:g3935162"
                     /translation="MVGKSKSPEFNFEDGMAYFDAVKVALQDTEPEKYQEFVRIFIDY
                     TANRFGFETLSASLQELLRDHANLCHGSNVLLPFRLQTAIPPEASAEFHINKVVGRSV
                     PPAVAVPTMDDATSYLNAVKEAFHDEPAKYMEITKLLTDLKARRINAASVIARMEELL
                     KDHLNLLLGFCVFLSPTRRFITKLKARFLGDGSQVVDSVLQILRMHSEGNKSKDEASQ
                     EVRALIQGHEDLLMELSEIFSDP"
     CDS             join(17912. .17988,18081. .18639,18726. .18875)
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.6"
                     /db_xref="PID:g3935163"
                     /translation="MEEVYHDEPEKFNHILHLIRDYFNHRDDRARITACMGELMRDHL
                     NLLVRFFDFFPAEASEGLAQLQTIAATSVSPDSTIDDAVSYINTVKEAFHDEPAKYYE
                     FFQLFYDIRYRYVLLCPSSILFALNFKSQSIQHCEVILFVIYYSLCRLIDVAGGITRV
                     EELLKAHKNLLVRLNAFLPPEAQRILHLKIEQRAASDINKRKRVASFIGKLKERFQGD
                     DRHVYESFLEILTMYQEGNKSVNDLYQEVGFLVGTLWLDYTFS"
     CDS             join(19430. .19507,19610. .19831,20016. .24392)
                     /note="putative polyprotein (AF039376)"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.7"
                     /db_xref="PID:g3935164"
                     /translation="MFINIVNRVIQVNCWSHTWHEASCLQVFDTKTQTCDPKPIPCSE
                     TFGGFDYKTLCLDGKFHVGPVPVDKTRIRLADYGGKRAVLWPNNIHSLRKMIWCAESF
                     KRSLNLNLAATLFFIIFHYFSFSSMASETVITSDSTSLLNVNMTNVTKLTSSNFLMWR
                     RQVQALLNGYDLTGYIDGSIVVPPATITANGAVTVNPAFKHWQRQDQLIYSALLGAIS
                     ISVQPILSRTTTSAEIWTKLMDTYAKPSWSHIQQLRQQIKQWKKDTKSIDEFFQGLVM
                     RFDQLALLGKPMESEEQMEVIVEGLSDDYKQVIDQIQGREVPPSLTEIHEKLLNHEVK
                     LQAAASSLPISANAASYRPPANNKHNNSNNYRGQNRNNNNRGANSYQQPRNDQPSSRG
                     YQGKCQICGVFGHSARRCSQLQMSGAYSTPSPSQYPNATVPWQPRANMAAMSYNPWLL
                     DSGATHHLTTDLNNLALHQPYNGGEEVTIADGSTLPITHTGSSTLSTQSRSLALNNIL
                     YVPNLHKNLISVYKLCNANKVSVEFFPAHFQVKDLSTGARLLQGRTKDELYEWPVPSN
                     TPISLFASPTPKTTLPSWHSRLGHPSPPVLKSLVSQFSLPVSNSSQKHFPCSHCLINK
                     SHKLPFYSNTIISYTPLEYVYSDVWTSPVTSVDNFKYYLILVDHYTRYTWLYPLKQKS
                     QVRETFVAFKALVENRFQTKIRTLYSDNGGEFIALRQFLLTHGISHLTSLPHTPEHNG
                     IAERKHRHILETGLTLLTQASIPTSYWTYAFGTAVYLINRLPSSVLNNESPYSKLFKT
                     SPNYLKLRVFGCSCFPWLRPYTNHKLERRSQPCVFLGYSLTQSAYLCLDRSSGRVYTS
                     RHVQFVEDQFPFSISDTHSVSNSSPEEASPSCHQPPSRIPIQSSSPPLVQAPSSLPPL
                     SSDSHRRPNAETSSSSSSTNNDVVVSKDNTQVDNRNNFIGPTSSSSAQSQNNSNPSSS
                     IQTQNEPNPSPSPTPQNSSPESSPSSSTSATSTVPNPPPPPPTNNHPMRTRAKNHITK
                     PKTKLSLLAKTVQTRPQIPNTVNQALRDEKWRNAMGEEINAQIRNNTFELVPPKPNQN
                     VISTKWIFTLKYLPNGTLDRYKARLVARGFRQQYGLHYSETFSPVVKSLTIRLVLQLA
                     VSRSWTIKQLDVNNAFLQGTLTDEVYVTQPPGFIDPDRPHHVCRLKKALYGLKQAPRA
                     WYQELRNFVCSLGFTNSLADTSVFVYINDIQIVYCLVYVDDIIVTGSSDALVMAFITA
                     LSRRFSLKDPTDLVYFLGIEATRTSQGLHLMQHKYVYDLLSRMKMLDAKPVSTPMATH
                     PKLSLYSGIALDEPGEYRTVIGSLQYLAFTRPDIAYAVNRLSQFMHRPTDIHWQAAKR
                     VLRYLAGTATHGILLRSNSPLSLHAFSDADWAGDNDDFVSTNAYIVYLGSTPIAWSSK
                     KQKGVARSSTEAEYRAVANTTSEIRWVCSLLTELGITLPKMPVIYCDNVGATYLSANP
                     VFHSRMKHLALDYHFIRDNVSAGALRVSHISTHDQLADALTKPLPRQHFLQFSSKIGV
                     SKLPPS"
     CDS             join(25673. .26084,26445. .26461)
                     /note="unknown;  similar to ESTs gb|T42386, gb|H36247,
                     gb|N38423, gb|T21830, and gb|AA585725"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.8"
                     /db_xref="PID:g3935165"
                     /translation="MMRRQRDEQQSRVFYDLSSLVLNLLRSPPMPISLPDHFPDPPAR
                     IRPPFSSHRFSSSSPSLAHISPSGFASLLLGISVALMLCGSVTFFIGFLLMPWVLALI
                     LVFYVAGIVSAISMVGRSILCYVLTPPSPSGKEISEWKLL"
     CDS             join(27209. .27349,27618. .28079)
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F17L21.9"
                     /db_xref="PID:g3935166"
                     /translation="MESSEDVENLSRAIEKLLHEKRKREASGDAFIEDADDQLFLSSL
```

ISQLESPNPKGKEDVITGEEEEKEEESADSSPSKGKSEGQRQLEESIEEIAKDIKKVK
KQNKITHVLLSATIILTLVWQLSEYSMIFMLKDRISHPVRSIGGMLNGMFKGKLRPIK
NQLAGTSNSNDQNNHGNGSHTGPQLQVPELLREFGFDDEE"

CDS complement(join(28500. .28571,29379. .29675))
/note="similar to nuclear transport factor 2; similar to ESTs gb|T21347, gb|N96312, and emb|F20149"
/codon_start=1
/evidence=not_experimental
/product="F17L21.10"
/db_xref="PID:g3935167"
/translation="MDPDAVAKAFVEHYYSTFDANRPGLVSLYQEGSMLTFEGQKIQG
SQNIVAKLTGLPFQQCKHNITTVDCQPSGPAGGMLVFVSGNLQLAGEQHALKFSQMFH
LISNQGNYYVFNDIFRLNYA"

CDS join(31665. .32171,32453. .32734,32937. .33106,33195. .33690, 33784. .33835,33950. .34179,34268. .34366,34453. .35558, 35648. .35827,35899. .35986,36081. .36149)
/note="similar to sensory protein barA (sp|P26607); similar to histidine kinase A (gi|1136289); similar to ESTs gb|AA712891 and gb|AA042438"
/codon_start=1
/product="F17L21.11"
/db_xref="PID:g3935168"
/translation="MSLFHVLGFGVKIGHLFWMLCCWFVSWFVDNGIEDKSGLLVGSV
GDLEKTKMTTLKKKNKMWFWNKISSSGLKIPSFSYQFLGSVKFNKAWWRKLVVVWVVF
WVLVSIWTFWYFSSQAMEKRKETLASMCDERARMLQDQFNVSMNHVQAMSILISTFHH
GKIPSAIDQRTFSEYTDRTSFERPLTSGVAYAMRVLHSEREEFERQQGWTIRKMYSLE
QNPVHKDDYDLEALEPSPVQEEYAPVIFAQDTVSHVVSLDMLSGKEDRENVLRARSSG
KGVLTAPFPLIKTNRLGVILTFAVYKRDLPSNATPKERIEATNGYLGGVFDIESLVEN
LLQQLASKQTILVNVYDITNHSQPISMYGTNVSADGLERVSPLIFGDPLRKHEMRCRY
LQLAHTYVCNFFLFARIQVLTFCCELLPLCRFKQKPPWPVLSMVTSFGILVIALLVAH
IIHATVSRIHKVEEDCDKMKQLKKKAEAADVAKSQFLATVSHEIRTPMNGVLGMLHML
MDTELDVTQQDYVRTAQASGKALVSLINEVLDQAKIESGKLELEEVRFDLRGILDDVL
SLFSSKSQQKGVELAVYISDRVPDMLIGDPGRFRQILTNLMGNSIKFTEKGHIFVTVH
LVDELFESIDGETASSPESTLSGLPVADRQRSWENFKAFSSNGHRSFEPSPPDINLIV
SVEDTGVGIPVEAQSRIFTPFMQVGPSISRTHGGTGIGLSISKCLVGLMKGEIGFSST
PKVGSTFTFTAVFSNGMQPAERKNDNNQPIFSEFRGMKAVVVDHRPARAKVSWYHFQR
LGIRVEVVPRVEQALHYLKIGTTTVNMILIEQEIWNREADDFIKKLQKDPLFLSPKLI
LLANSVESSISEALCTGIDPPIVIVKPLRASMLAATLQRGLGIGIREPPQHKGPPALI
LRNLLLGRKILIVDDNNVNLRVAAGALKKYGADVVCAESGIKAISLLKPPHEFDACFM
DIQMPEMDGFEATRRIRDMEEEMNKRIKNGEALIVENGNKTSWHLPVLAMTADVIQAT
HEECLKCGMDGYRIIRGDGAGRRTAKETQKLVSPMHLEVVEKQIDDCKKSVYRKAGVT
RPRT"

CDS join(36949. .36951,37408. .37564,37737. .37783)
/note="unknown; similar to EST gb|AA650671 and gb|T20610"
/codon_start=1
/evidence=not_experimental
/product="F17L21.12"
/db_xref="PID:g3935169"
/translation="MTTSKRLADRKIEKFDKNILKRGFVPETTTKKGKDYPVGPILLG
FFVFVVIGSSLFQIIRTATSGGMA"

CDS join(39626. .39777,40491. .41742)
/note="similar to fim protein; similar to ESTs gb|T42445, gb|T76780, gb|AA650733, and emb|Z17748"
/codon_start=1
/evidence=not_experimental
/product="F17L21.13"
/db_xref="PID:g3935170"
/translation="MEEELAMLRQLIGQLQELLHNGSPPPPSSSSSLSSSSPSFLVLH
HPQYQNGWCLPCIEDTSADDCCDIVMAGGKRPGIFKMLETVKPPVKRTRKERTQGKSC
TEVDEISGNMDQEIWQEFPQDLFEDVVSRLPMATFFQFRAVCRKWNALIDSDSFSRCF
TELPQTIPWFYTITHENVNSGQVYDPSLKKWHHPIIPALPKKSIVLPMASAGGLVCFL
DIGHRNFYVSNPLTKSFRELPARSFKVWSRVAVGMTLNGNSTSHGYKVLWVGCEGEYE
VYDSLSNVWTKRGTIPSNIKLPVLLNFKSQPVAIHSTLYFMLTDPEGILSYDMVSGKW
KQFIIPGPPDLSDHTLAACGERLMLVGLLTKNAATCVCIWELQKMTLLWKEVDRMPNI
WCLEFYGKHIRMNCLGNKGCLILLSLRSRQMNRLITYNAVTREWTKVPGCTVPRGRKR
LWIACGTAFHPSPTARA"

CDS complement(join(42083. .42239,42509. .42669,43186. .43188))
/note="unknown; similar to ESTs gb|T20610 and gb|AA586199"
/codon_start=1
/evidence=not_experimental
/product="F17L21.14"
/db_xref="PID:g3935171"
/translation="MTTSKRLADRKIEKFDKNILKRGFVPETTTKKGKDYPVGPILLG
FFVFVVIGSCESLGLQLAEAWHKPKHSQARRSSFQNVESSSVSFLFLSLYCRMCQLSV
YYIM"

CDS join(46024. .46685,46846. .46967,47100. .47351,47455. .47600)
/note="putative squamosa-promoter binding protein (AC002561)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.15"
/db_xref="PID:g3935172"
/translation="MDCNMVSSSQWDWEHLIMSNPSRTEDDSKQLPTEWEIEKGEGIE
SIVPHFSGLERVSSGSATSFWHTAVSKSSQSTSINSSSPEAKRCKLASESSPGDSCSN
IDFVQVKAPTALEVSVASAESDLCLKLGKRTYSEEYWGRNNNEISAVSMKLLTPSVVA
GKSKLCGQSMPVPRCQIDGCELDLSSAKGYHRKHKVCEKHSKCPKVSVSGLERRFCQQ
CSRFHAVSEFDEKKRSCRKRLSHHNARRRKPQGVFSMNPERVYDRRQHTNMLWNGVSL
NARSEEMYEWGNNTYDTKPRQTEKSFTLSFQRGNGSEDQLVASSSRMFSTSQTSGGFP
AGKSKFQLHGEDVGEYSGVLHESQDIHRALSLLSTSSDPLAQPHVQPFSLLCSYDVVP
K"

CDS complement(join(49339. .49493,49603. .49851,49967. .50088, 50218. .50882))
/note="similar to squamosa-promoter binding protein (gi|2673911)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.16"
/db_xref="PID:g3935173"
/translation="MDCNMVSSFPWDWENLIMSNQSKTENEKKQQSTEWEFEKGEGIE
SIVPDFLGFEKVSSGSATSFWHTAVSKSSQSTSINSSSPEDKRCNLASQSSPGDSSSN
IDFLQVKPSTALEVPIASAESDLCLKLGKRTYSEEFWGRNNNDLSAVSMNLLTPSVVA
RKKTKSCGQSMQVPRCQIDGCELDLSSSKDYHRKHRVCETHSKCPKVVVSGLERRFCQ
QCSRFHAVSEFDEKKRSCRKRLSHHNARRRKPQGVFPLNSERVFDRRQHTSMLWNGLS
LNTRSEEKYTWGTTYETKPTQMESGFTLSFQRGNGSEDQLFTGSTLSFSAFQTSGGFS
AGKSNIQLPDKGVGECSGGLHESHDFYSALSLLSTTSDSQGIKHTPVAEPPPIFGTFP
SHFI"

CDS complement(join(52788. .52954,53423. .53556,53832. .53974))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.17"
/db_xref="PID:g3935174"
/translation="MKDRMERLLILPFSLGCSTQSCVAVATTHQQKKPNQLIKSTFLL
TFLLVESDAKGSFRKDYTKMENNNNGANISDGIYRIIRSFKSFSHFFIRYEEETKERE
AEMEIGFPTDVKHLSHIGVDGTMTTFDNTSSSFPFSGFHLTGTVV"

CDS complement(join(55014. .55070,55246. .55329,55703. .55816, 56001. .56104,56456. .56705))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.18"
/db_xref="PID:g3935175"
/translation="MACKTILRSVFVSESRRTSGARRCFFLPPSPASVPVHGLFPAPK
SLCFNGFASVPERATRLNCSHNDQSDQGPPQEAVLKAISEVSKTDGRVGKTTNMIIGG
TVADDSAKDWLELDQKVNTYPTERGFTAIGTGGNDFVHAMVVAVESVIDRQIPEDCVK
QTLSSKGKYVSVNIGPIQVVSSEQVQAVYNAMRRDERMKYFL"

CDS complement(join(57354. .57485,57663. .57761,57936. .57983, 58075. .58244,58331. .58409,58699. .58797))
/note="similar to TOM20 (X92491); similar to ESTs gb|T444475, emb|Z26777, and emb|Z26778"
/codon_start=1
/product="F17L21.19"
/db_xref="PID:g3935176"
/translation="MEFSTADFERLIMFEHARKNSEAQYKNDPLDSENLLKWGGALLE
LSQFQPIPEAKLMLNDAISKLEEALTINPGKHQALWCIANAYTAHAFYVHDPEEAKEH
FDKATEYFQRAENEDPGNDTYRKSLDSSLKAPELHMQFMNQGMGQQILGGGGGGGGGG
MASSNQSSKKKKRNTEFTYDVCGWIILACGIVAWVGMAKSLGPPPPAR"

CDS join(59115. .59117,59214. .59250,59369. .59409,59889. .60119, 60204. .60464,60563. .60610)
/note="similar to 60S ribosomal protein L17-2 (sp|P35267); similar to ESTs gb|L33542 and gb|AA660016"
/codon_start=1
/evidence=not_experimental
/product="F17L21.20"
/db_xref="PID:g3935177"

```
                /translation="MVKYSQEPDNITKSCKARGADLRVHFKNTRETAHAIRKLPLNKA
                KRYLEDVIAHKQAIPFTRFCRGVGRTAQAKNRHSNGQGRWPAKSAQFVLDLLKNAESN
                AEVKGLDVDALFISHIQVNQAAKQRRRTYRAHGRINRMSSFLFDCFYNVCFFIHIVLT
                LDLILSYAAYMSNPCHIELILSEKEEPVKKEPETQLAAKSKKGASS"
CDS             join(61111. .61402,61486. .61688,61772. .61879,61963. .62628)
                /note="hypothetical protein"
                /codon_start=1
                /evidence=not_experimental
                /product="F17L21.21"
                /db_xref="PID:g3935178"
                /translation="MESGLISVDRWRNGSQAYFLTHIHSDHTRGLSGGWSQGPLYCSR
                TTASLFPSRFPGFDLSLLRVVPLFSWTSLSLRSPSSGSTVRLHLMAIDAHHCPGSIMF
                LFRGDFGCFLYTGDFRWDSDASDEARTTLVAAIDEFPVDILYLDNTYCNPIYSFPSRL
                VAVQLVADIIASHPSHDIIIAVDSLGKEDLLVHVSRILNIKIWVWPERLRTMHLLGFQ
                DIFTTDTSLTRVRAVPRYSFSIQTLEGLNTMCPTIGIMPSGLPWVKRPFKGDDKLSGS
                FLTASMKNETVSAKKELEAAAVHKFHDYMYSVHYSDHSCYEEIGEFIKLVKPKSMKGI
                VVSSSSYVDPLYYFGRICGANQPPQVLLMRPDIAEEFRAVRIKSYSASDKTRMLAKEK
                RWKRDSHSSLKRNKKRARIQVKCAKILEVD"
CDS             join(62993. .63429,63730. .63953,64026. .64405)
                /note="similar to MIPP proteins (gi|2191178, sp|P28575)"
                /codon_start=1
                /evidence=not_experimental
                /product="F17L21.22"
                /db_xref="PID:g3935179"
                /translation="MVVESSSSPIIPGLTDDVAELCVSKIPRSSFQITSQVCRRWRSF
                LRSQHFAAVRKLTGTVEEFLCVLMESECGRDVYWEVFDASGNKLGQIPPVPGPLKRGF
                GVAVLDGGKIVFFGGYTEVEGSGINSTTVSASADVYEFDPANNSWRKLAGMNIPRYNF
                AFAEVNGLLYVIRGYSTDTYSLSNAEVYNPKTNQWSLMHCPNRPVWRGFAFAFSSKLY
                AVGNGSRFIDIYDPKTQTWEELNSEQSVSVYSYTVVRNKVYFMDRNMPGRLGVFDPEE
                NSWSSVFVPPREGGFWVRLGVWNNKVLLFSRVCGHETLMYDLDKEKGSKWRVCDQIKP
                SASQLASVLINF"
CDS             complement(join(64930. .65189,65274. .65411,65541. .67868,
                68016. .68726,68873. .68971,69070. .69160,69275. .70105,
                70275. .70345,70747. .70813))
                /note="unknown; similar to ESTs gb|N96209, gb|R87018,
                gb|H76168, and gb|AA395332"
                /codon_start=1
                /evidence=not_experimental
                /product="F17L21.23"
                /db_xref="PID:g3935180"
                /translation="MAEGKFDLPDDLIFSKSSDQLKELASDNSIPLSPQWLYTKSSEY
                KMDVRSPTPVPMGNPSDPNPKDAWRLDAPEDKKDWKKIVHENETSRRWREEERETGLL
                GARKVDRRKTERRIDSVSSRETGDIKNAAASDRWNDVNSRAAVHEPRRDNKWSSRWGP
                DDKEKEARCEKVDINKDKEEPQSESQSVVSNVRATSERDSDTRDKWRPRHRMESQSGG
                PSSYRAAPGFGLDRGRAEGPNLGFTVGRGRASTIGRGSSTSLIGAGSALSPVFRYPRG
                KLLDMYRKQKPDSSLGRILTEMDEVASITQVALIEPLAFIAPDAEEEANLNGIWKGRI
                ISSEVYTSSGEESLGGNSLLKCRIPESGETKVDGALLGFMNGDNGSMKNNDSGLLGSH
                NGGLGAASSVPRLNSVASESYGSGGAGYQLSHGSPEAVRSVFTKSSVLDGSESVVGSF
                EQAYTGKLQQPDTEVDHSEGAMPPEEFLFLYIDPQGVIQGPFIGSDIISWFEQGFFGT
                DLQVRLASAPEGTPFQDLGRVMSYIKAESVHAHISDQKSELEETSLKANSEAGGSVAH
                VAESNDSSSLTGISRSFSVYNNPSGQDNFQRKSESEVYGRPPHAEDQSFLDFSAQDEE
                IVFPGRARVSGYASSVKSSTSMHDALMEFSGHSDIPVEVTTAATRNQNENKLHPFGVL
                WSELEGGSTPVNPLPNRSSGAMGEPSCSIENRPINSRRNSQIDPNISLDALSGNRMSQ
                FEHESNFFNHGDQLPSNQHHQQHFQNRDMLSHLHIGDQDLEHLITLQLQQQQKIQMQQ
                QQKIQLQQQQKIQLQQHQLEQEHQLHQKLLQEQQQSHARQLHFQQILQGQTPDTRFGQ
                SHDFPRSNSVDQMLLEQQMLNELQKSSGHPSQNFAPYIEQHAAGNFGRFTHEGHQREL
                LEQLFSTQMQSQYGQKQSQYGQMQSQHGQLQSEPIRSLEYQLLQQEQLMQLANGVRHN
                TLLEEQRHIDPLWPSDHSDQLLRTHPGIHRSHSSAGFRPLDFHQQQQRPHFEDQFSQL
                ERNRSYQQQLRLELLEHGLPFERSASGLNLDAVNGLGLSQGLELRDATAHMQSSGRLG
                NSTPGFSHQNPRIPLGESHFSHLEPTEGRWSGADTQLAGDWAESQFRRSNMDTEHDKM
                RSEIRRLGEDPNSWMVGGSTDDKSKQLFMELLHQRPGHQSAESPNMNRGYPYDRMVPS
                GLTPGIQTLGGLSDHGGNQNVSSAFGDRSFSDEQVNRVPGYGNNMGSLHHNSSLLSGI
                IDAGRSTQNETQAFSNMFGMNKDANDINTWNNVPPKNEGMGRMMSYDAQDRMGKQAVL
                DSLIQEELPVGTPGQQSSFNISGMLSFRVIASSKIFCYFSNFMVWKVDFYFVIMKAYP
                DILLLQTDTVTTWLEKIEGRIVPSHGQNSVLLKRPPSSHSSSSHEGLLERMSDTASRA
                AASSYSGIEGGVRRESGAAGNKGSTSEAASFSEMLKKSNSMKKVAAESTDATEGSKGG
                GGKKKGKKGRQIDPALLGFKVTSNRILMGEIHRADDF"
CDS             complement(join(73150. .73663,73910. .74151,74405. .74674))
                /note="similar to multiple exostoses type II protein
                EXT2.I (U72263); similar to ESTs dbj|D39982, gb|L37635,
                and dbj|C28418"
                /codon_start=1
                /evidence=not_experimental
                /product="F17L21.24"
                /db_xref="PID:g3935181"
                /translation="MFAAEIFMHRFLLSSPVRTRNPDEADWFYTPIYPTCDLTPTGLP
                LPFKSPRMMRSSIQLISSNWPYWNRTEGADHFFVVPHDFGACFHYQEEKAIERGILPL
                LQRATLVQTFGQRNHVCLDEGSITIPPFAPPQKMQAHFIPPDIPRSIFVYFRGLFYDV
                NNDPEGGYYARGARAAVWENFKNNPLFDISTDHPTTYYEDMQRAIFCLCPLGWAPWSP
                RLVEAVVFGCIPVIIADDIVLPFADAIPWEEIGVFVAEKDVPELDTILTSIPTEVILR
                KQRLLANPSMKRAMLFPQPAQPGDAFHQILNGLARKLPHDKSIYLKTGEKALNWTAGP
                VADLKPW"
CDS             join(76306. .76387,76478. .76588,76756. .76848,77250. .77399,
                77484. .77542,77636. .77692)
                /note="putative adenine phosphoribosyltransferase 1
                (sp|P31166); similar to ESTs gb|N65531, gb|R90631,
                gb|T21275, and gb|AA713070"
                /codon_start=1
                /evidence=not_experimental
                /product="F17L21.25"
                /db_xref="PID:g3935182"
                /translation="MATEDVQDPRIAKIASSIRVIPDFPKPGIMFQDITTLLLDTEAF
                KDTIALFVDRYKDKGISVVAGVEARGFIFGPPIALAIGAKFVPMRKPKKLPGKVISEE
                YSLEYGTDTIEMHVGAVEPGERAIIIDDLIATGGTLAAAIRLLERVGVKIVECACVIE
                LPELKGKEKLGETSLFVLVKSAA"
CDS             join(78862. .79059,79153. .79411,79507. .80678,80747. .80891,
                80999. .81309)
                /note="hypothetical protein"
                /codon_start=1
                /evidence=not_experimental
                /product="F17L21.26"
                /db_xref="PID:g3935183"
                /translation="MLCACSGEQFRFEDQPGSPESLATRDFSASGLSSRNGGGDWDSK
                LEDIQVDEAESTLKEALSLNYEEARALLGRLEYQRGNFDAALQVFKGIDIKVLTPRII
                KAIVEKTLPCKPRSKAVIVPPTTMSMHSVSLLLEAILLKARSLEELGSYKEAAEECKI
                ILDVVENALPSGMPDGISGFAKLQDIFQKALELLPLLWKKAGNHHETIASYRRALSRP
                WNLDPQRLAVTQKSLALVLLYGSVEACPKDNIEEAIVLLMLLVKKMVVGDIQWDPELM
                DHLTYALSMTGQFEVLANYLEQTLPGVYTRGERWYLLSLCYSAAGIDKAAINLLKMAL
                GPSESRQIPHIPLLLFGAKLCSKDPKHSRDGINFAHRLLDLGNSQSEHLLSQAHKFLG
                VCYGNAARSSKLDSERVFLQKKSLFSLNEAAKRGKADPELDVIFNLSVENAVQRNVQA
                ALDGAVEYSSMVGGVSTKGWKHLAIVLSAEKRLKDAESILDFTMEEAGDIEKIELLRL
                KAVLQMAQEQPKKAMKTCSSLLGLIRAQEKSEQSESLLQKFETEAWQDLASVYGKLGS
                WSDAETCLEKARSMCYYSPRGWNETGLCLEAKSLHEEALISFFLSLSIEPDHVPSIVS
                IAEVMMKSGDESLPTAKSFLMNALRLDPRNHDAWMKLGHVAKKQGLSQQAAEFYQAAY
                ELELSAPVQSFI"
CDS             complement(join(81833. .82924,84388. .84580,84662. .84922,
                85075. .85256,85340. .85917,86264. .86411,86652. .86814,
                86888. .87001,87154. .87217,87303. .87594,87699. .87842,
                87945. .88103))
                /note="hypothetical protein"
                /codon_start=1
                /evidence=not_experimental
                /product="F17L21.27"
                /db_xref="PID:g3935184"
                /translation="MFEYRCSSIDWKPSPVVALATSADGSQVAAAREDGSLELWLVSP
                GASSWHCQLIIHGDPKSRISHLAWFGTGSKGSSSARLFSSSIDGSISEWDLFDLKQKV
                VLDSIGFSIWQIALAPISIEAESKEAKLIQNGYLSEKSDDEEESGVEDDSELDELDEK
                VEALDRHLAAACDDGCVRIYYISESDKLTYYRSLPRVTGRALSVTWSADAQRIFSGSS
                DGLIRCWDANLCHEVYRITVGLGGLGNGSELCIWSLLSLRYGVLVSGDSTGSVQFWDS
                QHGTLLQAHSNHKGDVNALAASPSHNRVFSAGADGQVILYKLSGGTSRSQDFKPSSAQ
                KWDYIGCVRAHTHDIRALTVAVPISWEGSIPDRNAKVTSPKQRRKEPAGFSYHKWAHL
                GVPMLISAGDDAKLYAYSVQEFTKFPPHDICPAPQRVPMQMVHNTVFNQTSLLLVQDS
                CSLDILRIHISNDSSGRVSTKPLVRVKSKDARKIICSAISNTGSLFAYSDQVRPSLFE
                LKKNKIGKNPWSANRKRLPNLPFAHSMVFSCDCSRLIIAGHDRRIYIVDVGSMELLHS
                FTPRQEGQEGESPPHEPPITKLYTSSNDDWLAAINCFGDIYVFNLETQRQHWFISRLD
                GASVAAAGFHPRDNNVIVISTSSNQVFALDVEARELSKWSLLQTFCLPKSYQNFPGEV
                VGLSFSPSPCSSSVIIYSSRAKCLIEFGKPAEQDEDTDTPCNLSEKLEGKLASISMKL
                GNGAQKRRLEEYQKESKSNKRKKFEMFFFFIILDVMIKILNPHSHHSQTTTLKTAEIL
                SKYRPIAPKPGTPRVNDDDPSSSMSHKISQSPYLRNLWPQLQARPTRTRKRGRGGMGP
                SSLAMKRPKSSCGSSSTSTISTQRVLGPIKTLSFQAFTHHRLPNLPQVGYGFENGVSS
                TLVTLPLLQCSPPSSKCMEPEIKGKGVIDLNKTAEVIQERDFLTQLQGPITTTTTATT
                SRVISPQPIRPVCSKINVAYINPLTNPSPTSQTSKKSPREVEEDVESDDLPSVITDSN
                SRVRLVNSAYKEMMGQPECSWLDSMVRGKRICGEVMINFCESKIPVMTENNGFSCWVR
                IDWGRDGKEEYMHAFCDVTKLACDSKDYVFTWRFHTTTDRR"
CDS             join(88492. .88892,89156. .90053)
                /note="similar to lecithin:cholesterol acyltransferase
```

precursor (M26268)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.28"
/db_xref="PID:g3935185"
/translation="MKKISSHYSVVIAILVVVTMTSMCQAVGSNVYPLILVPGNGGNQ
LEVRLDREYKPSSVWCSSWLYPIHKKSGGWFRLWFDAAVLLSPFTRCFSDRMMLYYDP
DLDDYQNAPGVQTRVPHFGSTKSLLYLDPRLRDATSYMEHLVKALEKKCGYVNDQTIL
GAPYDFRYGLAASGHPSRVASQFLQDLKQLVEKTSSENEGKPVILLSHSLGGLFVLHF
LNRTTPSWRRKYIKHFVALAAPWGGTISQMKTFASGNTLGVPLVNPLLVRRHQRTSES
NQWLLPSTKVFHDRTKPLVVTPQVNYTAYEMDRFFADIGFSQGVVPYKTRVLPLTEEL
MTPGVPVTCIYGRGVDTPEVLMYGKGGFDKQPEIKYGDGDGTVNLASLAALKVDSLNT
VEIDGVSHTSILKDEIALKEIMKQISIINYELANVNAVNE"
CDS complement(join(90258. .90471,90595. .90775,90859. .90958))
/note="similar to MYB transcription factor isolog
(AC002335)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.29"
/db_xref="PID:g3935186"
/translation="MEWSLPVDLQEEILSRVPAKSLARWKSTPKQWKGPISIEFLHLL
RSTLKFPLLNLHLKSKSIYVTSFTAMVFCYAPPSTKDSRFGIHVQGKPSGFLADEQNK
VVVCCKWRSDSINTIYFVGENKHVQVDQRRGDLTLGQSCSFLMNYVPSFVQIQQGTLL
APRT"
CDS complement(join(95514. .95715,95789. .97565,98312. .98357,
99010. .99108))
/note="similar to kinesin light chain (Z25664); similar
to EST gb|R90451"
/codon_start=1
/evidence=not_experimental
/product="F17L21.30"
/db_xref="PID:g3935187"
/translation="MQLASYRGQGKIADFGFKKPRWVEGKLLKLNGKCRPYITIGIEN
RNRPYSCLFEIKKMEGGSVNESHSNADQMFDTTIEELCKNLCELQSSNQSPSRQSFGS
YGDESKIDSDLQHLALGEMRDIDILEDEGDEDEVAKPEEFDVKSNSSNLDLEVMPRDM
EKQTGKKNVTKSNVGVGGMRKKKVGGTKLQNGNEEPSSENVELARFLLNQARNLVSSG
DSTHKALELTHRAAKLFEASAENGKPCLEWIMCLHVTAAVHCKLKEYNEAIPVLQRSV
EIPVVEEGEEHALAKFAGLMQLGDTYAMVGQLESSISCYTEGLNIQKKVLGENDPRVG
ETCRYLAEALVQALRFDEAQQVCETALSIHRESGLPGSIAEAADRRLMGLICETKGDH
ENALEHLVLASMAMAANGQESEVAFVDTSIGDSYLSLSRFDEAICAYQKSLTALKTAK
GENHPAVGSVYIRLADLYNRTGKVREAKSYCENALRIYESHNLEISPEEIASGLTDIS
VICESMNEVEQAITLLQKALKIYADSPGQKIMIAGIEAQMGVLYYMMGKYMESYNTFK
SAISKLRATGKKQSTFFGIALNQMGLACIQLDAIEEAVELFEEAKCILEQECGPYHPE
TLGLYSNLAGAYDAIGRLDDAIKLLGHVVGVREEKLGTANPVTEDEKRRLAQLLKEAG
NVTGRKAKSLKTLIDSDLTSSSALR"
BASE COUNT 31054 a 18594 c 18732 g 31309 t
ORIGIN

Query Match 3.2%; Score 41; DB 39; Length 99689;
Best Local Similarity 49.5%; Pred. No. 3.9;
Matches 136; Conservative 0; Mismatches 135; Indels 4; Gaps 1;

```
Qy    662 ACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAAG 721
          ||| | |  | | ||||  |    | ||   |  | ||  || || || |   ||| ||
Db  12114 ACAAACCAGAAAATTTACTATGGGTATAAGGTTCACCACTAGAACTTGAATGATTTTAAT 12055

Qy    722 ACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAAT----CCA 777
            |  |||   |      |  ||  | | || ||||| |||     || ||||     ||
Db  12054 TTGAGCTGTATAGGTCTTTATATATATATAAAAAAAAATAGACCTTTTGCAATTGTATCA 11995

Qy    778 TAACATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATT 837
          |||  |   |    | | ||||| |||  ||  |   |||| |||  |    |  || | |
Db  11994 TAATACTAAATATAATAAACACATGTTAGTTGTAAATAAAAATTTCACATTAATTACAAT 11935

Qy    838 TGTTAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTAC 897
          |   | || | | |||| | |   | |  | |  ||   ||| || ||  |||||
Db  11934 TTACAGATTTTTATGTAAATTTTAATTATACATATATTTATGTAAATTTAGTAATAAATA 11875

Qy    898 ATGGAGACAGGTCAACCAAAGAGGGAGCTAGGCAA 932
          ||  | |  |   | |   | | | | ||  | ||
Db  11874 ATTTAAATTGTAAAGCTTTACAAGCATCTGTGAAA 11840
```

RESULT 14
AC000348/c
LOCUS AC000348 111390 bp DNA PLN 24-JUN-1997
DEFINITION Genomic sequence for Arabidopsis thaliana BAC T7N9, complete sequence.
ACCESSION AC000348
NID g2182287
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryotae; mitochondrial eukaryotes; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; vascular plants; seed plants; Magnoliophyta; Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 111390)
AUTHORS Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Shinn,P., Sun,H., Conway,A., Conway,A., Kurtz,D., Oji,O., Osborne,B., Shen,Y.K., Toriumi,M., Vysotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.
TITLE Genomic sequence for Arabidopsis thaliana BAC T7N9
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (24-APR-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 5 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 6 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (24-JUN-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT On Jun 10, 1997 this sequence version replaced gi:2155227.
FEATURES Location/Qualifiers
source 1. .111390
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
CDS join(1465. .1811,1909. .2068,2232. .2362,2562. .3125)
/note="similar to protein kinase, 5' incomplete; similar to EST gb|T14833"
/codon_start=3
/evidence=not_experimental
/product="T7N9.1"
/db_xref="PID:g2213581"
/translation="SSRISAKRSQSSRLSSLTIQSSSYNDDTSVASLQTPRSEGELLA
SPTLKAFTFNELKTATRNFRPDSVIGEGGFGYVYKGWIDERTLSPSKPGSGMVVAVKK
LKEEGFQGHRQWLAEVDCLGRLHHMNLVKLIGYCSKGDHIRLLVYEYMPKGSLENHLF
RRKKTCLSSGAEPIPWRTRIKVAIGAARGLAFLHEAQVIYRDFKASNILLDSEFNAKL
SDFGLAKVGPTGDRTHVSTQVMGTQGYAAPEYVATGRITAKSDVYSFGVVLLELLSGR
LTVDKTKVGVERNLVDWAIPYLGDKRKVFRIMDTKLGGQYPHKGACLTANTALQCLNQ
EPKLRPKMSDVLSTLEELEMTLKSGSISNSVMKLTSSSSSFTAKQRVRTPVADPVLSS
RRCRRVR"

CDS complement(join(9987. .11012,11109. .12706,12764. .13970, 14075. .14464))
/note="TA1-like retrotransposon"
/codon_start=1
/evidence=not_experimental
/product="T7N9.2"
/db_xref="PID:g2213582"
/translation="MVKKGGKMKTRGSTPDDEGSAVRSVREAAILEDQKLTTLKQQLA
QRSDVYGGNGFKVSDSGENPLLLHSSDHPGLSIVAHILDGSNYNSWSIAMRISLDAKN
KLGFVDGSLLRPSVDDSTFRIWSRCNSMVNNLPRRYQLEQAVMTLQQGKLDLSTYFTK
KKTLWEQLANTKSRSVKKCDCDQVKELLEEAETSRVIQFLMGLSDDFNTIRSQIFNMK
PRPGLNEIYNMLDQDESQRLVGFAAKSVPSPSPAAFQTQGVLNDQNTILLAQGNFKKP
KCTHCNRIGHTVDKCYKVHGYPPGHPRAKENTYVGSTNLASTDQIETQAPPTMSATGH
ETMSNDHIQQLISYLSTKLQSPSITSCFDKAIASSSNPVPSISQITGTFLSLYDSTYY
EMLTSSIPIETELSLRAWVIDSGASHHVTHERNLYHTYKALDRTFVRLPNGHTVKIEG
TGFIQLTDALSLHNVLFIPEFKFNLLSVSVLTKTLQSKVSFTSDECMIQALTKELMLG
KGSQVGNLYILNLDKSLVDVSSFPGHPSFAKIDTLSDVLMLPKQKINKDSSHCHVCHL
SKQKHLPFKSVNHIREKAFELVHIDTWGPFSVPTVDSYRYFLTIVDDFSRATWIYLLK
QKSDVLTVFPSFLKMVETQYHTKVCSVRSDNAHELKFNELFAKEGIKADHPCPETPEQ
NFVVERKHQHLLNVARALMFQSGIPLEYWGDCVLTAVFLINRLLSPVINNETPYERLT
KGKPDYSSLKAFGCLCYCSTSPKSRTKFDPRAKACIFLGYPMGYKGYKLLDIETYSVS
ISRHVIFYEDIFPFASSNITDAAKDFFPHIYLPAPNNDEHLPLVQSSSDAPHNHDESS
SMIFVPSEPKSTRQRKLPSHLQDFHCYNNTPTTTKTSPYPLTNYISYSYLSEPFGAFI
NIITATKLPQKYSEARLDKVWNDAMGKEISAFVRTGTWSICDLPAGKVAVGCKWIITI
KFLADGSIERHKARLVAKGYTQQEGIDFFNTFSPVAKMVTVKVLLSLAPKMKWYLHQL
DISNALLNGDLEEEIYMKLPPGYSEIQGQEVSPNAKCHGDHTLFVKAQDGFFLVVLVY
VDDILIASTTEAASAELTSQLSSFFQLRDLGEPKFFLGIEIARNADGISLCQRKYVLD
LLASSDFSDCKPSSIPMEPNQKLSKDTGTLLEDGKQYRRILGKLQYLCLTRPDINFAV
SKLAQYSSAPTDIHLQALHKILRYLKGTIGQGLFYGADTNFDLRGFSDSDWQTCPDTR
RCVTGFAIFVGNSLVSWRSKKQDVVSMSSAEAEYRAMSVATKELIWLGYILTAFKIPF
THPAYLYCDNEAALHIANNSVFHERTKHIENDCHKVRECIEAGILKTIFVRTDNQLAD
TLTKPLYPKPFRENNSKLGLLNIYEAQA"
CDS join(15527. .15581,15687. .15748,15834. .15911,15991. .16066, 16164. .16234,16347. .16448,16539. .16628,16756. .16830, 17067. .17099,17201. .17272,17475. .17525,17658. .17773, 17834. .17927)
/note="unknown; similar to ESTs gb|R87028, gb|W43320, gb|AA395193, and gb|T75928"
/codon_start=1
/evidence=not_experimental
/product="T7N9.3"
/db_xref="PID:g2213583"
/translation="MAMQAGVGLSRIFLLAGAGYTGTIMMKNGKLSDLLGELQSLVKG
MEKSGEGSEGDSDVSDAIAAQVRRLAMEIRQLASQQHITVMNGVSGANLQALAVPAAA
LGALGYGYMWWKGLSFTDLMYVTKANMAAAVANLTKNLEQVSETLAAAKRHLTQRIQN
LDDKVEKQIDLSKEINSQVISARENISSLEMDLESLHNLITGLDGKLDTLEYKQDVTN
VFMLNLYNYFGGKSTKLPEMLPVNQRARNLLADVETKGLKNLAEELFKSNGTQVTTTV
KQISLSKVNVKSRPLLSRCDKGNVNAGLLQLSVDMEVFLMAVWLAGCI"
CDS join(20094. .20132,20210. .20362,20796. .21014,21105. .21220, 21317. .21480,21572. .21726,21798. .21878)
/note="similar to T7N9.5; similar to ESTs gb|T21578, gb|T46181, gb|H36893, gb|T46690, and gb|T46636"
/codon_start=1
/evidence=experimental
/product="T7N9.4"
/db_xref="PID:g2213584"
/translation="MGSLDLPYASSFKGGSETFLQNVFESILKTYLRKNPMAKTIWEL
VKSVDNEKISYDHFFFRTFKVDGYGIDSLASFFMDYGYKVGGRLDFPKKKVQVLWLSP
PDVHFPDNGYGIGNGPLPRLVIAELLVEELSPESQEIIRKYLKPEGGKQAVLSSTLGS
LIWEKPTSTDFNQLAKESEFAAWTLVYGYTMNHLAFAVHRLKHRFSDIKCVKEYFEEN
GFELNKDGGVLKVSEDSLLLQVSAMSEKLVVEFADGVTQIVPASYIEFVERLVLPQFK
DMPCDEIKEFHRREGLEQASAYHIMESTRFTA"
CDS join(23301. .23339,23416. .23568,23660. .23878,23967. .24082, 24559. .24722,24830. .24993,25085. .25171)
/note="similar to T7N9.4; similar to ESTs gb|T46181, gb|R29915, gb|H74776, gb|H36893, gb|T46690, and gb|T46636"
/codon_start=1
/evidence=not_experimental
/product="T7N9.5"
/db_xref="PID:g2213585"
/translation="MGSFDLPHSSSFKGESEIFLRNVFENILKTYLRKNPTAKTIWEL
VQSLDNEKICYDHFTFRTLKVDGYGIDSLSSFFMAYGYKIGGGLDFPKKKLRVLWFSP
PDVHVPNDGHGLGNGPLPRLVIAEVLVDELSPESQGIIRKYLKQEGGKQAVLSSTLGS
LIWEKPTWTDFKQLAKESEFAAWTLIHGYTMNHLAFAVHRFKHRFSDIKFVKQRLEEK
GFKLNSDGEILKVSQDGLLFQVSSISERLPVTFADGVTETIPASYIEFTQRQVLPEFK
DVPLDEVIDIKEFHRREAFELDNANHVMESTRFTTKF"
CDS complement(join(28352. .29284,29376. .29813,30850. .31067, 31195. .31297))
/note="similar to transport protein; similar to ESTs emb|Z17614, gb|T14067, and emb|Z17615"
/codon_start=1
/evidence=not_experimental
/product="T7N9.6"
/db_xref="PID:g2213586"
/translation="MHGDVSKWEGYADWRNKAALRGRHGGMLAASFVLAVEILENLAF
LANASNLVLYLKNFMHMSLARSSSEVTTFMATAFLLALLGGFLADAFFSTFVIFLISA
SIEFLGLILLTIQARRPSLMPPPCKSSAALRCEVVGGSKAAFLFVGLYLVSLGIGGIK
GSLPSHGAEQFDEGTPKGRKQRSTFFNYYVFCLSCGALVAVTFVVWIEDNKGWEWGFG
VSTISIFLSILVFLLGSRFYKNKIPRGSPLTTIFKVLLAASIVSCSSKTSSNHFTSRE
VQSEHEEKTPSQSLTNSLTCLNKAIEGKTHHIWLECTVQQVEDVKIVLKMLPIFGCTI
MLNCCLAQLSTYSVHQAATMNRKIVNFNVPSASLPVFPVVFMLILAPTYDHLIIPFAR
KVTKSEIGITHLQRIGVGLVLSIVAMAVAALVELKRKQVAREAGLLDSEETLPITFLW
IALQYLFLGSADLFTLAGLLEFFFTEAPSSMRSLATSLSWASLALGYYLSSVMVPIVN
RVTKSAGQSPWLGEKLNRNRLDLFYWLMCVLSVVNFLHYLFWAKRYKYISTGSIS"
CDS join(33352. .33469,33706. .34339,34852. .34978,35132. .35506, 35592. .35867)
/note="similar to leucine zipper protein; similar to ESTs gb|T21286 and gb|T04740"
/codon_start=1
/evidence=not_experimental
/product="T7N9.7"
/db_xref="PID:g2213587"
/translation="MENSDTDSEVFFWFQNQNQNHSHKFPSSCFPPSSHSAFYGSSSM
INTETATMDEEDVCESYMMREITKKRKLTPIQLRLLEESFEEEKRLEPDRKLWLAEKL
GLQPSQVAVWFQNRRARYKTKQLEHDCDSLKASYAKLKTDWDILFVQNQTLKSKVQFL
NRLTSHYFQESVQNFDDTFKQVDLLKEKLKMQENLETQSIERKRLGEEGSSVKSDNTQ
YSEEEGLENQYSFPELAVLGFYYDPTLTASNLRQEPLKVTCADQMTQIQISDVTEPAS
SAHVNSCRSSSNDTRQYKKIEVTQRSSSMSRKRDKPYTNRHTPARISKRRRPWAPSSS
EHDEIIDKPITKPPPPPALVVMGLPANCSVLELKSRFEIYGSISRIRIHKDGIGSVSY
RTAESAEAAIAGSHEPSFGISIDSKKLEVVWATDPLVKWKEGVTAGEGKERTSSFSSK
LLRPVMPLRKHGRSSRLASAIVNPRSDNTKGISGDGGISSPATTSEVKQRNIVTYDDI
V"
CDS join(36482. .36677,36759. .36911,36993. .37151,37225. .37383, 37476. .37607,37687. .38132)
/note="similar to Human p619"
/codon_start=1
/evidence=not_experimental
/product="T7N9.8"
/db_xref="PID:g2213588"
/translation="MAEHGETQADEEEEEQVWSWGAGTDGQLGTTKLQDELLPQLLSL
TSLPSISMLACGGAHVIALTSGGKVFTWGRGSSGQLGHGDILNITLPKLVSFFDDSVI
TQAAAGWSHSGFVSDSGCIFTCGNGSFGQLGHGDTLSLSTPAKVSHFNNDSVKMVACG
MRHSLVLFAGNQVCGFGSGKRGQLGFSSDRIKSVNLPCVVSGLKDVEVVRISANGDHS
AAISADGQFFSWGRGFCGGPDVHAPQSLPSPLSFREVAVGWNHALLLTVDGEVFKLGS
TLNKQPEKQQLQIDSSEALFEKVPDFDGVKVMQIAAGAEHSAAVTGKRGGAILLFLSS
LHSANIKQTTFECLAENGEVKTWGWGEHGQLGLGNTNDQTSPELVSLGSIDLRTKEIK
VYCGSGFTYAVKRKQELSPTSS"
CDS complement(join(38404. .38499,38806. .38953,39043. .39188, 39284. .39498,39579. .39756,39845. .40010,40104. .40453))
/note="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.9"
/db_xref="PID:g2213589"
/translation="MCNKRNLGWDSEGSKDLETEILEFMKNSEKPGMFPSKKDLIRSG
RFDLVERIVNQGGWLSMGWDLDEQEEKVRVNENVTPQDLHIEKQLPNCNSPEMDKTLN
HGDLDLSSNLSSSTEQVESRNDSGIEGILTRLEKERNLSLGISVRENGKSNGAMHDIS
PNGSVPWSSRIVTASEIQEVDGSRGSGEYAQSRYQGAKSVSGKPGLSDSPTSETWRTW
SMRRAGFTDEDFEAAEISSSGLTGVKKDDTKKDSGDSMNGKDRIASSSEDVNKTHIKH
RLQQLQSELSSVLHSLRSPPDKVVTSKDSETTAGNLENLSDDWEYKENEIIHAQNKLR
STRAKLAVLEGKMAMAIIDAQRIVREKQRRIDHASRALRLLRTASIVWPNSASEVLLT
GSFDGWSTQIKFIVDGQWKVDPLRPIVTSGGYENNLLIIS"
CDS join(42641. .42830,43041. .43258,43725. .44287,44375. .45248)
/note="similar to transport protein; similar to ESTs gb|T43913 and gb|L46413"
/codon_start=1

/evidence=not_experimental
/product="T7N9.10"
/db_xref="PID:g2213590"
/translation="MPSLFSFLIGSCVQHHANIMTHEHTKLLLRVMQFNNLRKYIVAI
EMESEIKLSDFKFIFSTVSSNETLEKLGSIGVSANFMLYLRNVFHMEPVEAFNVYYLW
MGLTNFAPLLGALISDAYIGRFKTIAYASLFSILGLMTVTLTACLPQLHPPPCNNPHP
DECDDPNKLQLGILFLGLGFLSIGSGGIRPCSIPFGVDQFDQRTEQGLKGVASFFNWY
YLTLTMVLIFSHTVVVYLQTVSWVIGFSIPTSLMACAVVLFFVGMRFYVYVKPEGSVF
SGIARVIVAARKKRDLKISLVDDGTEEYYEPPVKPGVLSKLPLTDQFKFLDKAAVILD
GDLTSEGVPANKWRLCSIQEVEEVKCLIRVVPVWSAGIISIVAMTTQATFMVFQATKM
DRHMGPHFEIPAASITVISYITIGIWVPIYEHLLVPFLWRMRKFRVTLLQRMGIGIVF
AILSMFTAGFVEGVRRTRATEMTQMSVFWLALPLILMGLCESFNFIGLIEFFNSQFPE
HMRSIANSLFPLSFAAANYLSSLLVTTVHKVSGTKDHPDWLNKDLDRGKLDYFYYLIA
VLGVVNLVYFWYCAHRYQYKAGSQIEDFNEEKSLLDIEPNQRHDQSPS"
CDS complement(join(45500. .45856,45940. .46014,46641. .46814,
46865. .47557))
/note="unknown; similar to ESTs gb|T44556, gb|U74100,
gb|L38131, gb|AA395480,"
/codon_start=1
/evidence=not_experimental
/product="T7N9.11"
/db_xref="PID:g2213591"
/translation="MAATASSEASEGPVMGLINKRLRALRKKYNRITQMEESISQGKT
LNKEQEEVLRSKPAVVILIDELEKIRAPLSAAVTEEISLATQLNRASSDQTTASEQKE
VTDIPQEVSGGDDGAKLEDLVNFLYFGSLFDVKSQNEFTSIMLTRTHERSCCLSYDYV
TDDATDLLGDRDLDSISQLWSLMVSRPVDSSLSHKNALERCVEHAKLWLANSEQPIES
NCNTSCKIKIPKSICDADAALREKLKKIMASDYFTTTPEMKAPVDVAAAAGNYTSYQV
PVDVEASGHYQQKEEDASNSKEVESVVNDQSQQDEHQKVELVTEGEVVHGQQEQGYTQ
VGGAENAKRDYVPRGSYQNQRGRRGARRGGGGYQNGRGGRGGGGGYQNGRYESYDQSG
GNGYQRNYYNNRGRGRGGGGGGGNGHSYNNYQDPNVAVAS"
CDS complement(join(49016. .50053,51910. .52533))
/note="unknown, similar to ESTs dbj|C22029, gb|T46021, and
gb|T04347, gb|N37850"
/codon_start=1
/evidence=not_experimental
/product="T7N9.12"
/db_xref="PID:g2213592"
/translation="MELFTKGNNVKLRSHLDKFLVADDDQETIRQSRKGDARRAVWTV
EPVVDKPNLIRLKSSHGTYLTASNKPLLLGMTGEKVTQTASFNKLMDWQTQWEPERDG
FQVKLKSWCGKWMRANGGTPPWRNSVTHDEPHTSKTKNWLIWDVITIDGSDLENMSDG
DESSVSSPVGSEFGSEPGSPVSARSTKSSISRFASLGLSTSPRWSSKPKSTASSFNQK
ETPVSVSAMEFFQKAKAIRMRNSHNKYLTADDDEETVTQNRNGSTKNARWTVEPVRDS
FHVIRLKSCYGKYLTASNERFLLGATGKKVIQLKLSRVDSSVEWEPLREGSKIKLRTR
SGNYLRGNGGLPPWRNSVTHDVPHLSATQDSISWDVDVVEILTDSEFKTESAKAPPPK
TTSPPPHRRPTSSPLSAESPRTSSSLSDRSDSDVSTFPKFLKHLWFCLKLQLYRFPVI
ENFDFDFDQSVESPPKSDGRTIYYHVADEEGHVEDETTVGYAFTFKGNSVAELTQTLR
EETCMEDAVVCTRSPLNGKLFPLRLQLPPNNGTLHVILLPSSASL"
CDS complement(join(54477. .54572,54654. .54754,54815. .54899,
54991. .55080,55175. .55226,55321. .55427,55516. .55623,
55710. .55839,55935. .56029,56119. .56262,56383. .56484,
56600. .56698,56843. .56978,57249. .57307))
/note="unknown; similar to EST gb|AA394730"
/codon_start=1
/evidence=not_experimental
/product="T7N9.13"
/db_xref="PID:g2213593"
/translation="MRSTPLQMIALLPLTLILTREKKVILEAPNYDNDCVLANILAAH
YLSSFDPVRARSYALAAESRLGKATLYEKAVFEAVSYLLSENMDDDVALELHSKLLKK
FPKDLLSWKRVETLCSYMGRHDLSLPLFRKILPQNEGQVYVNGMLAFCLIELGHLREA
EEAARKGCEINENDSWAHHALCHVLQTECRFKEAVKFMEEHSDSWDSCSSLRFSHNWW
HVAVCYLEGGSHISKVEEVYDHQMWKELEKDDAVARDVYTDALGLLLRLDTRGKLDDG
FQDRLEKLADSLTDKAMWYQDWLFDITTIWALSKVEKTSLAHELLEGLKSRTSAMNPK
KQKLMQKAILLAEAVYEYGKGNYEIALELLGLDFDAANYKVIGGSGLQMDVFNEIWYK
LLLLNGKSSTATNSAFTTPEPAAIEVLEKVTKQRDGAPFLWRLLEESYSMEGKEKAGI
SAGKKAKALESSYFKSA"
CDS join(62848. .63715,63803. .64056,64127. .64417,64491. .64619,
64694. .64867,64960. .65067,65160. .65369)
/note="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.14"
/db_xref="PID:g2213594"
/translation="MKKSKLDNSSSQIRFGLVQFLLVVLLFYFLCMSFEIPFIFRTGS
GSGSDDVSSSSFADALPRPMVVGGGSREANWVVGEEEEADPHRHFKDPGRVQLRLPER
KMREFKSVSEIFVNESFFDNGGFSDEFSIFHKTAKHAISMGRKMWDGLDSGLIKPDKA
PVKTRIEKCPDMVSVSESEFVNRSRILVLPCGLTLGSHITVVATPHWAHVEKDGDKTA
MVSQFMMELQGLKAVDGEDPPRILHFNPRIKGDWSGRPVIEQNTCYRMQWGSGLRCDG
RESSDDEEYGIDFFDGEVKCERWKRDDDDGGNNGDDFDESKKTWWLNRLMGRRKKMIT
HDWDYPFAEGKLFVLTLRAGMEGYHISVNGRHITSFPYRTGFVLEDATGLAVKGNIDV
HSVYAASLPSTNPSFAPQKHLEMQRIWKAPSLPQKPVELFIGILSAGNHFAERMAVRK
SWMQQKLVRSSKVVARFFVALHARKEVNVDLKKEAEYFGDIVIVPYMDHYDLVVLKTV
AICEYGVNTVAAKYVMKCDDDTFVRVDAVIQEAEKVKGRESLYIGNINFNHKPLRTGK
WAVTFEEWPEEYYPPYANGPGYILSYDVAKFIVDDFEQKRLRLFKMEDVSMGMWVEKF
NETRPVAVVHSLKFCQFGCIEDYFTAHYQSPRQMICMWDKLQRLGKPQCCNMR"
CDS join(67041. .67367,67700. .68056)
/note="similar to Bronze-2 protein; similar to ESTs
gb|H76491, gb|H36232, gb|H37163, gb|AA389825, gb|H37163,
emb|Z26546"
/codon_start=1
/evidence=not_experimental
/product="T7N9.15"
/db_xref="PID:g2213595"
/translation="MAQNDTVKLIGSWSSPYSLRARVALHLKSVKYEYLDEPDVLKEK
SELLLKSNPIHKKVPVLLHGDLSISESLNVVQYVDEAWPSVPSILPSDAYDRASARFW
AQYIDDKCFAAVDAVVGAKDDEGKMAAVGKLMECLAILEETFQKSSKGLGFFGGETIG
YLDIACSALLGPISVIEAFSGVKFLRQETTPGLIKWAERFRAHEAVKPYMPTVEEVVA
FAKQKFNVQ"
CDS join(69317. .69699,69740. .70175)
/note="similar to Bronze-2 protein; similar to ESTs
gb|H76491, gb|H36232, gb|H37163, gb|AA389825, gb|H37163,
emb|Z26546"
/codon_start=1
/evidence=not_experimental
/product="T7N9.16"
/db_xref="PID:g2213596"
/translation="MAQNDTVKLIGCSDDPFSIRPRVALHLKSIKYEYLEEPDDDLGE
KSQLLLKSNPIHKKTPVLIHGDLAICESLNIVQYLDEAWPSDPSILPSNAYDRASARF
WAQYIDDKVKQFRDSLLHALFCFLSKLVKSRVMCLKQCFEAANALTGANNDEERIAAT
GKLTECLAILEETFQKSSKGLGFFGGETIGYLDIACAALLGPISVIEMFSADKFVREE
TTPGLIQWAVRFRAHEAVRPYMPTVEEVTELVKQRIEEGFKRNFKSNVSTSEYE"
CDS join(71067. .71225,71312. .71410,71569. .71670,71750. .71893,
72009. .72103,72210. .72339,72415. .72519,72624. .72730,
72806. .72857,72941. .73030,73120. .73204,73290. .73354,
73441. .73527)
/note="unknown; similar to ESTs gb|T88591 and gb|AA394730"
/codon_start=1
/evidence=not_experimental
/product="T7N9.17"
/db_xref="PID:g2213597"
/translation="MKVLSYGRKRKVILEAPLYDKDCVLGSILAAHFLSSSDPSRANS
YVEAAASNLEQSTPYEKAVYEAVTYLISEDRDDDLAFEMHTKLLKRFPKDLASLKRAQ
LLSFYMGQPDPFLGLVQQVLPANQEESYIHGLLAFPLLELGRMEEAAAASRKGYEINK
EDAWAHHCLCHVLQHECRFKEAVEFMEALAGTWPSCSSFMYTHNWWHVALCYLEGGSP
MSKVEEIYDHHIWKELEKDDAVPPEVYLNALGLLIRLDVRDALDGFEDRLKNLAVRLT
NQANWYLEWHLDILIVWALAKVGETSRAHELLEGLKFRLSKKNKKKQQVMQKGIQLGE
AVYEYARGNYEKALELLGSEFNAIGYKIVGASDEQIDVFNEMWCQLLLKTGQSSTAKE
VIRERIKARDGIPFMWRLLEKSYSMEGNAEAESSARERAKKLESCYF"
misc_feature 73851. .74562
/note="similarity to valyl tRNA synthetase; predicted
non-functional because of reading frame shift"
CDS join(75036. .75508,76209. .76649,76746. .77862,77947. .78357,
78445. .80043,80124. .80678)
/note="similar to disease resistance kinase"
/codon_start=1
/evidence=not_experimental
/product="T7N9.18"
/db_xref="PID:g2213598"
/translation="MVKTGDVSDQRSRLEWDVFLSFQRDARHKFTERLYEVLVKEQVR
VWNNDDVERGNHELGASLVEAMEDSVALVVVLSPNYAKSHWCLEELAMLCDLKSSLGR
LVLPIFYEVEPCMLRKQNGPYEMDFEEHSKRFSEEKIQRWRRALNIIGNIPGFVYRLK
YDVFLSFRGADTRDNFGDHLYKALKDKVRVFRDNEGMERGDEISSSLKAGMEDSAASV
IVISRNYSGSRWCLDELAMLCKMKSSLDRRILPIFYHVDPSHVRKQSDHIKKDFEEHQ
VRFSEEKEKVQEWREALTLVGNLAGYVCDKDSKDDDMIELVVKRVLAELSNTPEKVGE
FIVGLESPLKDLTGLIDTESSSGVQVLGLYGMGGIGKTTLAKAFYNKIVGNFEQRAFI
SDIRERSSAENGLVTLQKTLIKELFRLVPEIEDVSIGLEKIKANVHEKKIIVVLDDVD
HIDQVHALVGETRWYGQGTLIVITTRDSEILSKLSVNQQYEVKCLTEPQALKLFSYHS

LRKEEPTKNLLALSKKIVQISGLLPLAVEVFGSLLYDKKEEKDWQTQLDKLKKTQPGN
LQDVLELSFKSLDDEEKKVFLDIACLFLKMEIKKDEVVIVLKGCGLNAEAALSVLRQK
SLVKILANDTLWMHDQIRDMGRQMVLKESREDPGLRSRLWDRGEIMTVLNNMKGTSSI
RGIVLDFKKKFARDPTADEIVSRNLRNNPGIYSVFNYLKNKLVRFPAEEKPKSSEITI
PVESFAPMTKLRLLQINNVELEGNLKLLPSELKWIQWKGCPLENLPPDFLARQLSVLD
LSESGIRQVQTLRNKMVDENLKVVILRGCHSLEAIPDLSNHEALEKLVFEQCTLLVKV
PKSVGNLRKLIHLDFRRCSKLSEFLVDVSGLKLLEKLFLSGCSDLSVLPENIGAMTSL
KELLLDGTAIKNLPESINRLQNLEILSLRGCKIQELPLCIGTLKSLEKLYLDDTALKN
LPSSIGDLKNLQDLHLVRCTSLSKIPDSINELKSLKKLFINGSAVEELPLKPSSLPSL
YDFSAGDCKFLKQVPSSIGRLNSLLQLQLSSTPIEALPEEIGALHFIRELELRNCKFL
KFLPKSIGDMDTLYSLNLEGSNIEELPEEFGKLEKLVELRMSNCKMLKRLPESFGDLK
SLHRLYMKETLVSELPESFGNLSNLMVLEMLKKPLFRISESNVPGTSEEPRFVEVPNS
FSKLLKLEELDACSWRISGKIPDDLEKLSCLMKLNLGNNYFHSLPSSLVKLSNLQELS
LRDCRELKRLPPLPCKLEQLNLANCFSLESVSDLSELTILTDLNLTNCAKVVDIPGLE
HLTALKRLYMTGCNSNYSLAVKKRLSKASLKMMRNLSLPGNRVPDWFSQGPVTFSAQP
NRELRGVIIAVVVALNDETEDDDYQLPDVMEVQAQIHKLDHHKCTNTLHLSGVPRTNN
DQLHICRYSAFHPLVTMLKDGYTIQVIKRNPPIKQGVELKMHGIHLVYEGDDDLEGRE
NTLPETQQTVSQKLANFFSSFEEEEGETISESESTVI"
CDS join(81318. .81794,82755. .83236,83326. .84475,84562. .84972,
85047. .86648,86729. .87277)
/note="similar to disease resistance kinase"
/codon_start=1
/evidence=not_experimental
/product="T7N9.19"
/db_xref="PID:g2213599"
/translation="MMKTGAVVSNPRSRVKWDAFLSFQRDTSHNFTDRLYEALVKEEL
RVWNDDLERVDHDHDHELRPSLVEAIEDSVAFVVVLSPNYANSHLRLEELAKLCDLKC
LMVPIFYKVEPREVKEQNGPFEKDFEEHSKRFGEEKIQRWKGAMTTVGNISGFICGYE
IQLEMETGVVPNRLKYSVFLSFRGFDTRTNFCERLYIALNEKQNVRVFRDNEGMEKGD
KIDPSLFEAIEDSAASVIILSTNYANSSWCLDELALLCDLRSSLKRPMIPIFYGVNPE
DVRKQSGEFRKDFEEKAKSFDEETIQRWKRAMNLVGNIPGYVCTAKTVGDDNEGINRE
KVDDMIDLVVKKVVAAVRNRPEIVADYTVGLESPIKDLMKLFNTESSSGIQVMGLYGM
GGIGKTTLAKAFYNKIIVNFNRHRVFIESVRGKSSDQDGLVNLQKTLIKELFRLVPEI
EDVSIGLEKIKENVHEKKIIVVLDDVDHIDQVNALVGETSWYGEGSLIVITTRDSEIL
SKLSVNQQYEVKCLTEPQALKLFSFYSLRKEKPPTQGLLELSKKIAEVTGLLPLAVKV
FGSHFYDKDENEWQVELEKLKTQQDKLHGVLALSFKSLDEEEKKIFLDIACLFLKMDI
TKEEVVDILKGCGLNAEAALRVLIQKSLLTILTDDTLWMHDQIRDMGRQMVHKESSDD
PEMRSRLWDRGEIMNVLDYMKGTSSIRGIVLDFNKKFARDHTADEIFSSNLRNNPGIY
SVFNYLKNKLVRFPAEEKPKRSEITIPVESFAPMKKLRLLQINNVELEGDLKLLPSEL
KWIQWKGFPLENLPPDILSRQLGVLDLSESGVRRVKTLPRKRGDENLKVVNLRGCHGL
EAIPDLSNHNALEKLVLERCNLLVKVPRSVGNLGKLLQLDLRRCSSLSEFLGDVSGLK
CLEKFFLSGCSNLSVLPENIGSMPCLKELLLDGTAISNLPYSIFRLQKLEKLSLMGCR
SIEELPSCVGYLTSLEDLYLDDTALRNLPSSIGDLKNLQKLHLMRCTSLSTIPETINK
LMSLKELFINGSAVEELPIETGSLLCLTDLSAGDCKFLKQVPSSIGGLNSLLQLQLDS
TPIEALPEEIGDLHFIRQLDLRNCKSLKALPKTIGKMDTLYSLNLVGSNIEELPEEFG
KLENLVELRMNNCKMLKRLPKSFGDLKSLHRLYMQETLVAELPESFGNLSNLMVLEML
KKPLFRISESNVPGTSEEPRFVEVPNSFSKLLKLEELDACSWRISGKIPDDLEKLSCL
MKLNLGNNYFHSLPSSLVKLSNLQELSLRDCRELKRLPPLPCKLEQLNLANCFSLESV
SDLSELTILTDLNLTNCAKVVDIPGLEHLTALKRLYMTGCNSNYSLAVKKRLSKASLK
MMRNLSLPGNRVPDWFSQGPVTFSAQPNRELRGVIIAVVVALNDETEDDDYQLPDVME
VQAQIHKLDHHKCTNTLHLSGVPRTNNDQLHICRYSAFHPLVTMLKDGYTIQVIKRNP
PIKQGVELKMHGIHLVYEGDDDLEGRENTLPEAQQTVSQKLANFFSSFEEGEASAKGD
STVA"
CDS complement(88382. .90187)
/note="similar to S-receptor kinase; similar to ESTs
gb|T20510, gb|T45307, gb|R30033, gb|Z26201, gb|AA394682.
gb|Z26201, and gb|T43096"
/codon_start=1
/evidence=not_experimental
/product="T7N9.20"
/db_xref="PID:g2213600"
/translation="MKKIFITLLWLLFISSFLCSSSSAEDDVLCLQGLKNSLIDPSSR
LSSWSFPNSSASSICKLTGVSCWNEKENRIISLQLQSMQLAGEIPESLKLCRSLQSLD
LSGNDLSGSIPSQICSWLPYLVTLDLSGNKLGGSIPTQIVECKFLNALILSDNKLSGS
IPSQLSRLDRLRRLSLAGNDLSGTIPSELARFGGDDFSGNNGLCGKPLSRCGALNGRN
LSIIIVAGVLGAVGSLCVGLVIFWWFFIREGSRKKKGYGAGKSKDDSDWIGLLRSHKL
VQVTLFQKPIVKIKLGDLMAATNNFSSGNIDVSSRTGVSYKADLPDGSALAVKRLSAC
GFGEKQFRSEMNKLGELRHPNLVPLLGYCVVEDERLLVYKHMVNGTLFSQLHNGGLCD
AVLDWPTRRAIGVGAAKGLAWLHHGCQPPYLHQFISSNVILLDDDFDARITDYGLAKL
VGSRDSNDSSFNNGDLGELGYVAPEYSSTMVASLKGDVYGFGIVLLELVTGQKPLSVI
NGVEGFKGSLVDWVSQYLGTGRSKDAIDRSICDKGHDEEILQFLKIACSCVVSRPKER
PTMIQVYESLKNMADKHGVSEHYDEFPLVFNKQEA"
CDS complement(91271. .92998)
/note="unknown, similar to ESTs emb|F14134, gb|T21189,
gb|T43096, gb|T1189, gb|Z37659"
/codon_start=1
/evidence=not_experimental
/product="T7N9.21"
/db_xref="PID:g2213601"
/translation="MTEYENGKKRKVRNKQQVKVQFLSQRYLILCFCCFFVLLFFLSS
DRISTLSVRSDSLRPSLRVPTLSVLSSSMDSFHRGRFPPLSVEDRVQFPDHLLLILSH
GIGKGEKNLVCVYRGVKEETLVLPSISSDEFDEFRSIVRCPNAPLNYSSSVDLQFRGD
LVKKKMKKQSRRVHNWEKVGYEAVIDGDTVVVFVKGLTRRPHKESDPSYYKCQFEIEN
SEEKEVTQAIAAAQEVVRCGLPESLKLNPEMMFRVSVIHIDPRGRTTPALPSVARIYG
SDSIEKKEKKSGVKHELCVCTMLWNQAPFLREWIMYHSWLGVERWFIYDNNSDDGIQE
EIELLSSENYNVSRHVWPWIKTQEAGFSHCAVRAKEECNWVGFFDVDEFYYFPTHRSQ
GLPSKNALKSLVSNYTSWDLVGEIRTDCHSYGPSGLTSVPSQGVTVGYTCRQANPERH
KSIIRPELLTSSLLNEVHHFQLKEGVGHMSLVESVAVVNHYKYQVWDTFKAKFYRRVA
TYVVDWQENQNQGSKDRAPGLGTEAIEPPDWKRRFCEVWDTGLKDLVMSNFADQVTGY
LPWQRQQQE"
CDS complement(join(97226. .97606,97721. .98247,98334. .98433,
98516. .99376))
/note="unknown; similar to EST gb|N96069"
/codon_start=1
/evidence=not_experimental
/product="T7N9.22"
/db_xref="PID:g2213602"
/translation="MSVHGRFPASPPISLSPSSSSTSPSSQSPSTPPDLKQRVIACLN
KLADRDTLALASAELDSIARNLTHDSFSPFLNCIHNTDSSVKSPVRKQCVALLSVLSR
YHGDSLTPHLAKMVSTVIRRLRDPDSSVRSACAVATADMSAHVTRQPFASVAKPLIET
LIQEGDSNLQIGAALCLAASVDAATDPESEQLRKSLPKIGKLLKSDGFKAKAALLSAV
GSIITAGGAGTKPVLDWLVPVLIEFLSSEDWAARKSAAEALGKVATAEDLASQYKKTC
TTALESRRFDKVKSVRETMNRALNLWKEVSTDDEASLSPSRSSTDDGNIGCFSSVTRS
STIDVGLKSARPKKVTPIMKRSPSLPVNRSYAATRQKENLPKRNQGNMTMLVEEASSV
DNKGPHFTPVKKSSEETEEKANSGGPDIIKHTISEKSREDSKVSSFGGLRSGSRVAPC
SDDGDSVKNCKDDVEESKKDSEELSLIREQLALIENQQSSLLDLLQKFMGTSQSGIQS
LESRVSGLEMALDEISCDLAVSNGRVPRNSSGCAGDSCSKLPGTEFLSPKFWRKTEER
PRNRNTANEMAAYDQGMRESTDTNNGQRGGSVFQKRSRRDQFQDCMHTTLQKPTTR"
CDS join(105265. .105404,105539. .105663,106614. .106903)
/note="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.23"
/db_xref="PID:g2213603"
/translation="MAGEGSKPKATVDDAYAYIRTVKSTFHNDPDKYDDFMAIMKNFK
ARKIDRNTCIEEVKELLKGHRDLISGFNAFLPKCLEIADWYNIEVLEAELQALLMAMQ
HTWSKGYRKIIFEGDNRTVESLLNGNTRNFELHNLIIEIQHWRKKFSNAIIKWSFRST
NKAADRLAKGELENNVTFVSHSTR"
CDS join(107696. .107835,107952. .108198,108274. .108375,
108464. .108535)
/note="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.24"
/db_xref="PID:g2213604"
/translation="MVERRVQVEPTLSDAHSYITAVKEAFHDEPTKYEEFIKLMNDIR
DHGVDKASGIAKLTELIKGHPRLLRGLSFFFPQVNRDIHHEAKRTIILKDKATIPPEA
AYRGAKSTYTKIKQIEPDWENFMNMLKTRFRSLDTHVVESFLKIMIMYDEGKKSEKEV
QEEVVDLLYYHEDLIDKFFRLFNMRK"
CDS join(110027. .110187,110323. .110453,110972. .111093)
/note="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.25"
/db_xref="PID:g2213605"
/translation="MVGGGSKHVGEGSKPRATKDDAYAYLRAVRAKFHNDSKKYDDFV
AVMTNFKARRIDRDGCIKEVEQLLKGHRDLISGFNAFLPKCLEIKNYYFGAGEDLEVV
YTAKRKLLARSVTKRSSSSSSSSVYFLFRSCEII"
BASE COUNT 35883 a 19310 c 20289 g 35908 t
ORIGIN

Query Match 3.2%; Score 41; DB 8; Length 111390;
Best Local Similarity 49.5%; Pred. No. 3.9;
Matches 136; Conservative 0; Mismatches 135; Indels 4; Gaps 1;

```
Qy    662 ACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAAG 721
          ||| | |  | | ||||  |    | ||   |  | ||  || || || |   ||| ||
Db 109688 ACAAACCAGAAAATTTACTATGGGTATAAGGTTCACCACTAGAACTTGAATGATTTTAAT 109629

Qy    722 ACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAAT----CCA 777
            |  |||   |      |  ||  | | || ||||| |||     || ||||     ||
Db 109628 TTGAGCTGTATAGGTCTTTATATATATATAAAAAAAAATAGACCTTTTGCAATTGTATCA 109569

Qy    778 TAACATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATT 837
          ||| |   |    | | ||||| |||  ||  |   |||| |||  |    |  || | |
Db 109568 TAATACTAAATATAATAAACACATGTTAGTTGTAAATAAAAATTTCACATTAATTACAAT 109509

Qy    838 TGTTAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTAC 897
          |   | || | | |||| | |   | |  | |  ||   ||| || ||  |||||
Db 109508 TTACAGATTTTTATGTAAATTTTAATTATACATATATTTATGTAAATTTAGTAATAAATA 109449

Qy    898 ATGGAGACAGGTCAACCAAAGAGGGAGCTAGGCAA 932
          ||  | |  |   | |   | | | | ||  | ||
Db 109448 ATTTAAATTGTAAAGCTTTACAAGCATCTGTGAAA 109414
```

RESULT 15
AC000348/c
LOCUS AC000348 111390 bp DNA PLN 24-JUN-1997
DEFINITION Genomic sequence for Arabidopsis thaliana BAC T7N9, complete sequence.
ACCESSION AC000348
NID g2182287
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryotae; mitochondrial eukaryotes; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; vascular plants; seed plants; Magnoliophyta; Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 111390)
AUTHORS Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Shinn,P., Sun,H., Conway,A., Conway,A., Kurtz,D., Oji,O., Osborne,B., Shen,Y.K., Toriumi,M., Vysotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.
TITLE Genomic sequence for Arabidopsis thaliana BAC T7N9
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (24-APR-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 5 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 6 (bases 1 to 111390)
AUTHORS Ecker,J.R.
TITLE Direct Submission
JOURNAL Submitted (24-JUN-1997) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT On Jun 10, 1997 this sequence version replaced gi:2155227.

FEATURES Location/Qualifiers
source 1. .111390
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
CDS join(1465. .1811,1909. .2068,2232. .2362,2562. .3125)
/note="similar to protein kinase, 5' incomplete; similar to EST gb|T14833"
/codon_start=3
/evidence=not_experimental
/product="T7N9.1"
/db_xref="PID:g2213581"
/translation="SSRISAKRSQSSRLSSLTIQSSSYNDDTSVASLQTPRSEGELLA
SPTLKAFTFNELKTATRNFRPDSVIGEGGFGYVYKGWIDERTLSPSKPGSGMVVAVKK
LKEEGFQGHRQWLAEVDCLGRLHHMNLVKLIGYCSKGDHIRLLVYEYMPKGSLENHLF
RRKKTCLSSGAEPIPWRTRIKVAIGAARGLAFLHEAQVIYRDFKASNILLDSEFNAKL
SDFGLAKVGPTGDRTHVSTQVMGTQGYAAPEYVATGRITAKSDVYSFGVVLLELLSGR
LTVDKTKVGVERNLVDWAIPYLGDKRKVFRIMDTKLGGQYPHKGACLTANTALQCLNQ
EPKLRPKMSDVLSTLEELEMTLKSGSISNSVMKLTSSSSSFTAKQRVRTPVADPVLSS
RRCRRVR"
CDS complement(join(9987. .11012,11109. .12706,12764. .13970, 14075. .14464))
/note="TA1-like retrotransposon"
/codon_start=1
/evidence=not_experimental
/product="T7N9.2"
/db_xref="PID:g2213582"
/translation="MVKKGGKMKTRGSTPDDEGSAVRSVREAAILEDQKLTTLKQQLA
QRSDVYGGNGFKVSDSGENPLLLHSSDHPGLSIVAHILDGSNYNSWSIAMRISLDAKN
KLGFVDGSLLRPSVDDSTFRIWSRCNSMVNNLPRRYQLEQAVMTLQQGKLDLSTYFTK
KKTLWEQLANTKSRSVKKCDCDQVKELLEEAETSRVIQFLMGLSDDFNTIRSQIFNMK
PRPGLNEIYNMLDQDESQRLVGFAAKSVPSPSPAAFQTQGVLNDQNTILLAQGNFKKP
KCTHCNRIGHTVDKCYKVHGYPPGHPRAKENTYVGSTNLASTDQIETQAPPTMSATGH
ETMSNDHIQQLISYLSTKLQSPSITSCFDKAIASSSNPVPSISQITGTFLSLYDSTYY
EMLTSSIPIETELSLRAWVIDSGASHHVTHERNLYHTYKALDRTFVRLPNGHTVKIEG
TGFIQLTDALSLHNVLFIPEFKFNLLSVSVLTKTLQSKVSFTSDECMIQALTKELMLG
KGSQVGNLYILNLDKSLVDVSSFPGHPSFAKIDTLSDVLMLPKQKINKDSSHCHVCHL
SKQKHLPFKSVNHIREKAFELVHIDTWGPFSVPTVDSYRYFLTIVDDFSRATWIYLLK
QKSDVLTVFPSFLKMVETQYHTKVCSVRSDNAHELKFNELFAKEGIKADHPCPETPEQ
NFVVERKHQHLLNVARALMFQSGIPLEYWGDCVLTAVFLINRLLSPVINNETPYERLT
KGKPDYSSLKAFGCLCYCSTSPKSRTKFDPRAKACIFLGYPMGYKGYKLLDIETYSVS
ISRHVIFYEDIFPFASSNITDAAKDFFPHIYLPAPNNDEHLPLVQSSSDAPHNHDESS
SMIFVPSEPKSTRQRKLPSHLQDFHCYNNTPTTTKTSPYPLTNYISYSYLSEPFGAFI
NIITATKLPQKYSEARLDKVWNDAMGKEISAFVRTGTWSICDLPAGKVAVGCKWIITI
KFLADGSIERHKARLVAKGYTQQEGIDFFNTFSPVAKMVTVKVLLSLAPKMKWYLHQL
DISNALLNGDLEEEIYMKLPPGYSEIQGQEVSPNAKCHGDHTLFVKAQDGFFLVVLVY
VDDILIASTTEAASAELTSQLSSFFQLRDLGEPKFFLGIEIARNADGISLCQRKYVLD
LLASSDFSDCKPSSIPMEPNQKLSKDTGTLLEDGKQYRRILGKLQYLCLTRPDINFAV
SKLAQYSSAPTDIHLQALHKILRYLKGTIGQGLFYGADTNFDLRGFSDSDWQTCPDTR
RCVTGFAIFVGNSLVSWRSKKQDVVSMSSAEAEYRAMSVATKELIWLGYILTAFKIPF
THPAYLYCDNEAALHIANNSVFHERTKHIENDCHKVRECIEAGILKTIFVRTDNQLAD
TLTKPLYPKPFRENNSKLGLLNIYEAQA"
CDS join(15527. .15581,15687. .15748,15834. .15911,15991. .16066, 16164. .16234,16347. .16448,16539. .16628,16756. .16830, 17067. .17099,17201. .17272,17475. .17525,17658. .17773, 17834. .17927)
/note="unknown; similar to ESTs gb|R87028, gb|W43320, gb|AA395193, and gb|T75928"
/codon_start=1
/evidence=not_experimental
/product="T7N9.3"
/db_xref="PID:g2213583"
/translation="MAMQAGVGLSRIFLLAGAGYTGTIMMKNGKLSDLLGELQSLVKG
MEKSGEGSEGDSDVSDAIAAQVRRLAMEIRQLASQQHITVMNGVSGANLQALAVPAAA
LGALGYGYMWWKGLSFTDLMYVTKANMAAAVANLTKNLEQVSETLAAAKRHLTQRIQN
LDDKVEKQIDLSKEINSQVISARENISSLEMDLESLHNLITGLDGKLDTLEYKQDVTN
VFMLNLYNYFGGKSTKLPEMLPVNQRARNLLADVETKGLKNLAEELFKSNGTQVTTTV
KQISLSKVNVKSRPLLSRCDKGNVNAGLLQLSVDMEVFLMAVWLAGCI"
CDS join(20094. .20132,20210. .20362,20796. .21014,21105. .21220, 21317. .21480,21572. .21726,21798. .21878)
/note="similar to T7N9.5; similar to ESTs gb|T21578, gb|T46181, gb|H36893, gb|T46690, and gb|T46636"
/codon_start=1

/evidence=experimental
/product="T7N9.4"
/db_xref="PID:g2213584"
/translation="MGSLDLPYASSFKGGSETFLQNVFESILKTYLRKNPMAKTIWEL
VKSVDNEKISYDHFFFRTFKVDGYGIDSLASFFMDYGYKVGGRLDFPKKKVQVLWLSP
PDVHFPDNGYGIGNGPLPRLVIAELLVEELSPESQEIIRKYLKPEGGKQAVLSSTLGS
LIWEKPTSTDFNQLAKESEFAAWTLVYGYTMNHLAFAVHRLKHRFSDIKCVKEYFEEN
GFELNKDGGVLKVSEDSLLLQVSAMSEKLVVEFADGVTQIVPASYIEFVERLVLPQFK
DMPCDEIKEFHRREGLEQASAYHIMESTRFTA"

CDS join(23301. .23339,23416. .23568,23660. .23878,23967. .24082,
24559. .24722,24830. .24993,25085. .25171)
/note="similar to T7N9.4; similar to ESTs gb|T46181,
gb|R29915, gb|H74776, gb|H36893, gb|T46690, and gb|T46636"
/codon_start=1
/evidence=not_experimental
/product="T7N9.5"
/db_xref="PID:g2213585"
/translation="MGSFDLPHSSSFKGESEIFLRNVFENILKTYLRKNPTAKTIWEL
VQSLDNEKICYDHFTFRTLKVDGYGIDSLSSFFMAYGYKIGGGLDFPKKKLRVLWFSP
PDVHVPNDGHGLGNGPLPRLVIAEVLVDELSPESQGIIRKYLKQEGGKQAVLSSTLGS
LIWEKPTWTDFKQLAKESEFAAWTLIHGYTMNHLAFAVHRFKHRFSDIKFVKQRLEEK
GFKLNSDGEILKVSQDGLLFQVSSISERLPVTFADGVTETIPASYIEFTQRQVLPEFK
DVPLDEVIDIKEFHRREAFELDNANHVMESTRFTTKF"

CDS complement(join(28352. .29284,29376. .29813,30850. .31067,
31195. .31297))
/note="similar to transport protein; similar to ESTs
emb|Z17614, gb|T14067, and emb|Z17615"
/codon_start=1
/evidence=not_experimental
/product="T7N9.6"
/db_xref="PID:g2213586"
/translation="MHGDVSKWEGYADWRNKAALRGRHGGMLAASFVLAVEILENLAF
LANASNLVLYLKNFMHMSLARSSSEVTTFMATAFLLALLGGFLADAFFSTFVIFLISA
SIEFLGLILLTIQARRPSLMPPPCKSSAALRCEVVGGSKAAFLFVGLYLVSLGIGGIK
GSLPSHGAEQFDEGTPKGRKQRSTFFNYYVFCLSCGALVAVTFVVWIEDNKGWEWGFG
VSTISIFLSILVFLLGSRFYKNKIPRGSPLTTIFKVLLAASIVSCSSKTSSNHFTSRE
VQSEHEEKTPSQSLTNSLTCLNKAIEGKTHHIWLECTVQQVEDVKIVLKMLPIFGCTI
MLNCCLAQLSTYSVHQAATMNRKIVNFNVPSASLPVFPVVFMLILAPTYDHLIIPFAR
KVTKSEIGITHLQRIGVGLVLSIVAMAVAALVELKRKQVAREAGLLDSEETLPITFLW
IALQYLFLGSADLFTLAGLLEFFFTEAPSSMRSLATSLSWASLALGYYLSSVMVPIVN
RVTKSAGQSPWLGEKLNRNRLDLFYWLMCVLSVVNFLHYLFWAKRYKYISTGSIS"

CDS join(33352. .33469,33706. .34339,34852. .34978,35132. .35506,
35592. .35867)
/note="similar to leucine zipper protein; similar to ESTs
gb|T21286 and gb|T04740"
/codon_start=1
/evidence=not_experimental
/product="T7N9.7"
/db_xref="PID:g2213587"
/translation="MENSDTDSEVFFWFQNQNQNHSHKFPSSCFPPSSHSAFYGSSSM
INTETATMDEEDVCESYMMREITKKRKLTPIQLRLLEESFEEEKRLEPDRKLWLAEKL
GLQPSQVAVWFQNRRARYKTKQLEHDCDSLKASYAKLKTDWDILFVQNQTLKSKVQFL
NRLTSHYFQESVQNFDDTFKQVDLLKEKLKMQENLETQSIERKRLGEEGSSVKSDNTQ
YSEEEGLENQYSFPELAVLGFYYDPTLTASNLRQEPLKVTCADQMTQIQISDVTEPAS
SAHVNSCRSSSNDTRQYKKIEVTQRSSSMSRKRDKPYTNRHTPARISKRRRPWAPSSS
EHDEIIDKPITKPPPPPALVVMGLPANCSVLELKSRFEIYGSISRIRIHKDGIGSVSY
RTAESAEAAIAGSHEPSFGISIDSKKLEVVWATDPLVKWKEGVTAGEGKERTSSFSSK
LLRPVMPLRKHGRSSRLASAIVNPRSDNTKGISGDGGISSPATTSEVKQRNIVTYDDI
V"

CDS join(36482. .36677,36759. .36911,36993. .37151,37225. .37383,
37476. .37607,37687. .38132)
/note="similar to Human p619"
/codon_start=1
/evidence=not_experimental
/product="T7N9.8"
/db_xref="PID:g2213588"
/translation="MAEHGETQADEEEEEQVWSWGAGTDGQLGTTKLQDELLPQLLSL
TSLPSISMLACGGAHVIALTSGGKVFTWGRGSSGQLGHGDILNITLPKLVSFFDDSVI
TQAAAGWSHSGFVSDSGCIFTCGNGSFGQLGHGDTLSLSTPAKVSHFNNDSVKMVACG
MRHSLVLFAGNQVCGFGSGKRGQLGFSSDRIKSVNLPCVVSGLKDVEVVRISANGDHS
AAISADGQFFSWGRGFCGGPDVHAPQSLPSPLSFREVAVGWNHALLLTVDGEVFKLGS
TLNKQPEKQQLQIDSSEALFEKVPDFDGVKVMQIAAGAEHSAAVTGKRGGAILLFLSS
LHSANIKQTTFECLAENGEVKTWGWGEHGQLGLGNTNDQTSPELVSLGSIDLRTKEIK
VYCGSGFTYAVKRKQELSPTSS"

CDS complement(join(38404. .38499,38806. .38953,39043. .39188,
39284. .39498,39579. .39756,39845. .40010,40104. .40453))
/note="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.9"
/db_xref="PID:g2213589"
/translation="MCNKRNLGWDSEGSKDLETEILEFMKNSEKPGMFPSKKDLIRSG
RFDLVERIVNQGGWLSMGWDLDEQEEKVRVNENVTPQDLHIEKQLPNCNSPEMDKTLN
HGDLDLSSNLSSSTEQVESRNDSGIEGILTRLEKERNLSLGISVRENGKSNGAMHDIS
PNGSVPWSSRIVTASEIQEVDGSRGSGEYAQSRYQGAKSVSGKPGLSDSPTSETWRTW
SMRRAGFTDEDFEAAEISSSGLTGVKKDDTKKDSGDSMNGKDRIASSSEDVNKTHIKH
RLQQLQSELSSVLHSLRSPPDKVVTSKDSETTAGNLENLSDDWEYKENEIIHAQNKLR
STRAKLAVLEGKMAMAIIDAQRIVREKQRRIDHASRALRLLRTASIVWPNSASEVLLT
GSFDGWSTQIKFIVDGQWKVDPLRPIVTSGGYENNLLIIS"

CDS join(42641. .42830,43041. .43258,43725. .44287,44375. .45248)
/note="similar to transport protein; similar to ESTs
gb|T43913 and gb|L46413"
/codon_start=1
/evidence=not_experimental
/product="T7N9.10"
/db_xref="PID:g2213590"
/translation="MPSLFSFLIGSCVQHHANIMTHEHTKLLLRVMQFNNLRKYIVAI
EMESEIKLSDFKFIFSTVSSNETLEKLGSIGVSANFMLYLRNVFHMEPVEAFNVYYLW
MGLTNFAPLLGALISDAYIGRFKTIAYASLFSILGLMTVTLTACLPQLHPPPCNNPHP
DECDDPNKLQLGILFLGLGFLSIGSGGIRPCSIPFGVDQFDQRTEQGLKGVASFFNWY
YLTLTMVLIFSHTVVVYLQTVSWVIGFSIPTSLMACAVVLFFVGMRFYVYVKPEGSVF
SGIARVIVAARKKRDLKISLVDDGTEEYYEPPVKPGVLSKLPLTDQFKFLDKAAVILD
GDLTSEGVPANKWRLCSIQEVEEVKCLIRVVPVWSAGIISIVAMTTQATFMVFQATKM
DRHMGPHFEIPAASITVISYITIGIWVPIYEHLLVPFLWRMRKFRVTLLQRMGIGIVF
AILSMFTAGFVEGVRRTRATEMTQMSVFWLALPLILMGLCESFNFIGLIEFFNSQFPE
HMRSIANSLFPLSFAAANYLSSLLVTTVHKVSGTKDHPDWLNKDLDRGKLDYFYYLIA
VLGVVNLVYFWYCAHRYQYKAGSQIEDFNEEKSLLDIEPNQRHDQSPS"

CDS complement(join(45500. .45856,45940. .46014,46641. .46814,
46865. .47557))
/note="unknown; similar to ESTs gb|T44556, gb|U74100,
gb|L38131, gb|AA395480,"
/codon_start=1
/evidence=not_experimental
/product="T7N9.11"
/db_xref="PID:g2213591"
/translation="MAATASSEASEGPVMGLINKRLRALRKKYNRITQMEESISQGKT
LNKEQEEVLRSKPAVVILIDELEKIRAPLSAAVTEEISLATQLNRASSDQTTASEQKE
VTDIPQEVSGGDDGAKLEDLVNFLYFGSLFDVKSQNEFTSIMLTRTHERSCCLSYDYV
TDDATDLLGDRDLDSISQLWSLMVSRPVDSSLSHKNALERCVEHAKLWLANSEQPIES
NCNTSCKIKIPKSICDADAALREKLKKIMASDYFTTTPEMKAPVDVAAAAGNYTSYQV
PVDVEASGHYQQKEEDASNSKEVESVVNDQSQQDEHQKVELVTEGEVVHGQQEQGYTQ
VGGAENAKRDYVPRGSYQNQRGRRGARRGGGGYQNGRGGRGGGGYQNGRYESYDQSG
GNGYQRNYYNNRGRGRGGGGGGGNGHSYNNYQDPNVAVAS"

CDS complement(join(49016. .50053,51910. .52533))
/note="unknown, similar to ESTs dbj|C22029, gb|T46021, and
gb|T04347, gb|N37850"
/codon_start=1
/evidence=not_experimental
/product="T7N9.12"
/db_xref="PID:g2213592"
/translation="MELFTKGNNVKLRSHLDKFLVADDDQETIRQSRKGDARRAVWTV
EPVVDKPNLIRLKSSHGTYLTASNKPLLLGMTGEKVTQTASFNKLMDWQTQWEPERDG
FQVKLKSWCGKWMRANGGTPPWRNSVTHDEPHTSKTKNWLIWDVITIDGSDLENMSDG
DESSVSSPVGSEFGSEPGSPVSARSTKSSISRFASLGLSTSPRWSSKPKSTASSFNQK
ETPVSVSAMEFFQKAKAIRMRNSHNKYLTADDDEETVTQNRNGSTKNARWTVEPVRDS
FHVIRLKSCYGKYLTASNERFLLGATGKKVIQLKLSRVDSSVEWEPLREGSKIKLRTR
SGNYLRGNGGLPPWRNSVTHDVPHLSATQDSISWDVDVVEILTDSEFKTESAKAPPPK
TTSPPPHRRPTSSPLSAESPRTSSSLSDRSDSDVSTFPKFLKHLWFCLKLQLYRFPVI
ENFDFDFDQSVESPPKSDGRTIYYHVADEEGHVEDETTVGYAFTFKGNSVAELTQTLR
EETCMEDAVVCTRSPLNGKLFPLRLQLPPNNGTLHVILLPSSASL"

CDS complement(join(54477. .54572,54654. .54754,54815. .54899,
54991. .55080,55175. .55226,55321. .55427,55516. .55623,
55710. .55839,55935. .56029,56119. .56262,56383. .56484,
56600. .56698,56843. .56978,57249. .57307))
/note="unknown; similar to EST gb|AA394730"
/codon_start=1

/evidence=not_experimental
/product="T7N9.13"
/db_xref="PID:g2213593"
/translation="MRSTPLQMIALLPLTLILTREKKVILEAPNYDNDCVLANILAAH
YLSSFDPVRARSYALAAESRLGKATLYEKAVFEAVSYLLSENMDDDVALELHSKLLKK
FPKDLLSWKRVETLCSYMGRHDLSLPLFRKILPQNEGQVYVNGMLAFCLIELGHLREA
EEAARKGCEINENDSWAHHALCHVLQTECRFKEAVKFMEEHSDSWDSCSSLRFSHNWW
HVAVCYLEGGSHISKVEEVYDHQMWKELEKDDAVARDVYTDALGLLLRLDTRGKLDDG
FQDRLEKLADSLTDKAMWYQDWLFDITTIWALSKVEKTSLAHELLEGLKSRTSAMNPK
KQKLMQKAILLAEAVYEYGKGNYEIALELLGLDFDAANYKVIGGSGLQMDVFNEIWYK
LLLLNGKSSTATNSAFTTPEPAAIEVLEKVTKQRDGAPFLWRLLEESYSMEGKEKAGI
SAGKKAKALESSYFKSA"

CDS join(62848. .63715,63803. .64056,64127. .64417,64491. .64619,
64694. .64867,64960. .65067,65160. .65369)
/note="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.14"
/db_xref="PID:g2213594"
/translation="MKKSKLDNSSSQIRFGLVQFLLVVLLFYFLCMSFEIPFIFRTGS
GSGSDDVSSSSFADALPRPMVVGGGSREANWVVGEEEEADPHRHFKDPGRVQLRLPER
KMREFKSVSEIFVNESFFDNGGFSDEFSIFHKTAKHAISMGRKMWDGLDSGLIKPDKA
PVKTRIEKCPDMVSVSESEFVNRSRILVLPCGLTLGSHITVVATPHWAHVEKDGDKTA
MVSQFMMELQGLKAVDGEDPPRILHFNPRIKGDWSGRPVIEQNTCYRMQWGSGLRCDG
RESSDDEEYGIDFFDGEVKCERWKRDDDDGGNNGDDFDESKKTWWLNRLMGRRKKMIT
HDWDYPFAEGKLFVLTLRAGMEGYHISVNGRHITSPPYRTGFVLEDATGLAVKGNIDV
HSVYAASLPSTNPSFAPQKHLEMQRIWKAPSLPQKPVELFIGILSAGNHFAERMAVRK
SWMQQKLVRSSKVVARFFVALHARKEVNVDLKKEAEYFGDIVIVPYMDHYDLVVLKTV
AICEYGVNTVAAKYVMKCDDDTFVRVDAVIQEAEKVKGRESLYIGNINFNHKPLRTGK
WAVTFEEWPEEYYPPYANGPGYILSYDVAKFIVDDFEQKRLRLFKMEDVSMGMWVEKF
NETRPVAVVHSLKFCQFGCIEDYFTAHYQSPRQMICMWDKLQRLGKPQCCNMR"

CDS join(67041. .67367,67700. .68056)
/note="similar to Bronze-2 protein; similar to ESTs
gb|H76491, gb|H36232, gb|H37163, gb|AA389825, gb|H37163,
emb|Z26546"
/codon_start=1
/evidence=not_experimental
/product="T7N9.15"
/db_xref="PID:g2213595"
/translation="MAQNDTVKLIGSWSSPYSLRARVALHLKSVKYEYLDEPDVLKEK
SELLLKSNPIHKKVPVLLHGDLSISESLNVVQYVDEAWPSVPSILPSDAYDRASARFW
AQYIDDKCFAAVDAVVGAKDDEGKMAAVGKLMECLAILEETFQKSSKGLGFFGGETIG
YLDIACSALLGPISVIEAFSGVKFLRQETTPGLIKWAERFRAHEAVKPYMPTVEEVVA
FAKQKFNVQ"

CDS join(69317. .69699,69740. .70175)
/note="similar to Bronze-2 protein; similar to ESTs
gb|H76491, gb|H36232, gb|H37163, gb|AA389825, gb|H37163,
emb|Z26546"
/codon_start=1
/evidence=not_experimental
/product="T7N9.16"
/db_xref="PID:g2213596"
/translation="MAQNDTVKLIGCSDDPFSIRPRVALHLKSIKYEYLEEPDDDLGE
KSQLLLKSNPIHKKTPVLIHGDLAICESLNIVQYLDEAWPSDPSILPSNAYDRASARF
WAQYIDDKVKQFRDSLLHALFCFLSKLVKSRVMCLKQCFEAANALTGANNDEERIAAT
GKLTECLAILEETFQKSSKGLGFFGGETIGYLDIACAALLGPISVIEMFSADKFVREE
TTPGLIQWAVRFRAHEAVRPYMPTVEEVTELVKQRIEEGFKRNFKSNVSTSEYE"

CDS join(71067. .71225,71312. .71410,71569. .71670,71750. .71893,
72009. .72103,72210. .72339,72415. .72519,72624. .72730,
72806. .72857,72941. .73030,73120. .73204,73290. .73354,
73441. .73527)
/note="unknown; similar to ESTs gb|T88591 and gb|AA394730"
/codon_start=1
/evidence=not_experimental
/product="T7N9.17"
/db_xref="PID:g2213597"
/translation="MKVLSYGRKRKVILEAPLYDKDCVLGSILAAHFLSSSDPSRANS
YVEAAASNLEQSTPYEKAVYEAVTYLISEDRDDDLAFEMHTKLLKRFPKDLASLKRAQ
LLSFYMGQPDPFLGLVQQVLPANQEESYIHGLLAFPLLELGRMEEAAAASRKGYEINK
EDAWAHHCLCHVLQHECRFKEAVEFMEALAGTWPSCSSFMYTHNWWHVALCYLEGGSP
MSKVEEIYDHHIWKELEKDDAVPPEVYLNALGLLIRLDVRDALDGFEDRLKNLAVRLT
NQANWYLEWHLDILIVWALAKVGETSRAHELLEGLKFRLSKKNKKKQQVMQKGIQLGE
AVYEYARGNYEKALELLGSEFNAIGYKIVGASDEQIDVFNEMWCQLLLKTGQSSTAKE

misc_feature VIRERIKARDGIPFMWRLLEKSYSMEGNAEAESSARERAKKLESCYF"
73851. .74562
/note="similarity to valyl tRNA synthetase; predicted
non-functional because of reading frame shift"

CDS join(75036. .75508,76209. .76649,76746. .77862,77947. .78357,
78445. .80043,80124. .80678)
/note="similar to disease resistance kinase"
/codon_start=1
/evidence=not_experimental
/product="T7N9.18"
/db_xref="PID:g2213598"
/translation="MVKTGDVSDQRSRLEWDVFLSFQRDARHKFTERLYEVLVKEQVR
VWNNDDVERGNHELGASLVEAMEDSVALVVVLSPNYAKSHWCLEELAMLCDLKSSLGR
LVLPIFYEVEPCMLRKQNGPYEMDFEEHSKRFSEEKIQRWRRALNIIGNIPGFVYRLK
YDVFLSFRGADTRDNFGDHLYKALKDKVRVFRDNEGMERGDEISSSLKAGMEDSAASV
IVISRNYSGSRWCLDELAMLCKMKSSLDRRILPIFYHVDPSHVRKQSDHIKKDFEEHQ
VRFSEEKEKVQEWREALTLVGNLAGYVCDKDSKDDDMIELVVKRVLAELSNTPEKVGE
FIVGLESPLKDLTGLIDTESSSGVQVLGLYGMGGIGKTTLAKAFYNKIVGNFEQRAFI
SDIRERSSAENGLVTLQKTLIKELFRLVPEIEDVSIGLEKIKANVHEKKIIVVLDDVD
HIDQVHALVGETRWYGQGTLIVITTRDSEILSKLSVNQQYEVKCLTEPQALKLFSYHS
LRKEEPTKNLLALSKKIVQISGLLPLAVEVFGSLLYDKKEEKDWQTQLDKLKKTQPGN
LQDVLELSFKSLDDEEKKVFLDIACLFLKMEIKKDEVVIVLKGCGLNAEAALSVLRQK
SLVKILANDTLWMHDQIRDMGRQMVLKESREDPGLRSRLWDRGEIMTVLNNMKGTSSI
RGIVLDFKKKFARDPTADEIVSRNLRNNPGIYSVFNYLKNKLVRFPAEEKPKSSEITI
PVESFAPMTKLRLLQINNVELEGNLKLLPSELKWIQWKGCPLENLPPDFLARQLSVLD
LSESGIRQVQTLRNKMVDENLKVVILRGCHSLEAIPDLSNHEALEKLVFEQCTLLVKV
PKSVGNLRKLIHLDFRRCSKLSEFLVDVSGLKLLEKLFLSGCSDLSVLPENIGAMTSL
KELLLDGTAIKNLPESINRLQNLEILSLRGCKIQELPLCIGTLKSLEKLYLDDTALKN
LPSSIGDLKNLQDLHLVRCTSLSKIPDSINELKSLKKLFINGSAVEELPLKPSSLPSL
YDFSAGDCKFLKQVPSSIGRLNSLLQLQLSSTPIEALPEEIGALHFIRELELRNCKFL
KFLPKSIGDMDTLYSLNLEGSNIEELPEEFGKLEKLVELRMSNCKMLKRLPESFGDLK
SLHRLYMKETLVSELPESFGNLSNLMVLEMLKKPLFRISESNVPGTSEEPRFVEVPNS
FSKLLKLEELDACSWRISGKIPDDLEKLSCLMKLNLGNNYFHSLPSSLVKLSNLQELS
LRDCRELKRLPPLPCKLEQLNLANCFSLESVSDLSELTILTDLNLTNCAKVVDIPGLE
HLTALKRLYMTGCNSNYSLAVKKRLSKASLKMMRNLSLPGNRVPDWFSQGPVTFSAQP
NRELRGVIIAVVVALNDETEDDDYQLPDVMEVQAQIHKLDHHKCTNTLHLSGVPRTNN
DQLHICRYSAFHPLVTMLKDGYTIQVIKRNPPIKQGVELKMHGIHLVYEGDDDLEGRE
NTLPETQQTVSQKLANFFSSFEEEEGETISESESTVI"

CDS join(81318. .81794,82755. .83236,83326. .84475,84562. .84972,
85047. .86648,86729. .87277)
/note="similar to disease resistance kinase"
/codon_start=1
/evidence=not_experimental
/product="T7N9.19"
/db_xref="PID:g2213599"
/translation="MMKTGAVVSNPRSRVKWDAFLSFQRDTSHNFTDRLYEALVKEEL
RVWNDDLERVDHDHDHELRPSLVEAIEDSVAFVVVLSPNYANSHLRLEELAKLCDLKC
LMVPIFYKVEPREVKEQNGPFEKDFEEHSKRFGEEKIQRWKGAMTTVGNISGFICGYE
IQLEMETGVVPNRLKYSVFLSFRGFDTRTNFCERLYIALNEKQNVRVFRDNEGMEKGD
KIDPSLFEAIEDSAASVIILSTNYANSSWCLDELALLCDLRSSLKRPMIPIFYGVNPE
DVRKQSGEFRKDFEEKAKSFDEETIQRWKRAMNLVGNIPGYVCTAKTVGDDNEGINRE
KVDDMIDLVVKKVVAAVRNRPEIVADYTVGLESPIKDLMKLFNTESSSGIQVMGLYGM
GGIGKTTLAKAFYNKIIVNFNRHRVFIESVRGKSSDQDGLVNLQKTLIKELFRLVPEI
EDVSIGLEKIKENVHEKKIIVVLDDVDHIDQVNALVGETSWYGEGSLIVITTRDSEIL
SKLSVNQQYEVKCLTEPQALKLFSFYSLRKEKPPTQGLLELSKKIAEVTGLLPLAVKV
FGSHFYDKDENEWQVELEKLKTQQDKLHGVLALSFKSLDEEEKKIFLDIACLFLKMDI
TKEEVVDILKGCGLNAEAALRVLIQKSLLTILTDDTLWMHDQIRDMGRQMVHKESSDD
PEMRSRLWDRGEIMNVLDYMKGTSSIRGIVLDFNKKFARDHTADEIFSSNLRNNPGIY
SVFNYLKNKLVRFPAEEKPKRSEITIPVESFAPMKKLRLLQINNVELEGDLKLLPSEL
KWIQWKGFPLENLPPDILSRQLGVLDLSESGVRRVKTLPRKRGDENLKVVNLRGCHGL
EAIPDLSNHNALEKLVLERCNLLVKVPRSVGNLGKLLQLDLRRCSSLSEFLGDVSGLK
CLEKFFLSGCSNLSVLPENIGSMPCLKELLLDGTAISNLPYSIFRLQKLEKLSLMGCR
SIEELPSCVGYLTSLEDLYLDDTALRNLPSSIGDLKNLQKLHLMRCTSLSTIPETINK
LMSLKELFINGSAVEELPIETGSLLCLTDLSAGDCKFLKQVPSSIGGLNSLLQLQLDS
TPIEALPEEIGDLHFIRQLDLRNCKSLKALPKTIGKMDTLYSLNLVGSNIEELPEEFG
KLENLVELRMNNCKMLKRLPKSFGDLKSLHRLYMQETLVAELPESFGNLSNLMVLEML
KKPLFRISESNVPGTSEEPRFVEVPNSFSKLLKLEELDACSWRISGKIPDDLEKLSCL
MKLNLGNNYFHSLPSSLVKLSNLQELSLRDCRELKRLPPLPCKLEQLNLANCFSLESV
SDLSELTILTDLNLTNCAKVVDIPGLEHLTALKRLYMTGCNSNYSLAVKKRLSKASLK
MMRNLSLPGNRVPDWFSQGPVTFSAQPNRELRGVIIAVVVALNDETEDDDYQLPDVME
VQAQIHKLDHHKCTNTLHLSGVPRTNNDQLHICRYSAFHPLVTMLKDGYTIQVIKRNP
PIKQGVELKMHGIHLVYEGDDDLEGRENTLPEAQQTVSQKLANFFSSFEEGEASAKGD
STVA"

```
CDS             complement(88382. .90187)
                /note="similar to S-receptor kinase; similar to ESTs
                gb|T20510, gb|T45307, gb|R30033, gb|Z26201, gb|AA394682.
                gb|Z26201, and gb|T43096"
                /codon_start=1
                /evidence=not_experimental
                /product="T7N9.20"
                /db_xref="PID:g2213600"
                /translation="MKKIFITLLWLLFISSFLCSSSSAEDDVLCLQGLKNSLIDPSSR
                LSSWSFPNSSASSICKLTGVSCWNEKENRIISLQLQSMQLAGEIPESLKLCRSLQSLD
                LSGNDLSGSIPSQICSWLPYLVTLDLSGNKLGGSIPTQIVECKFLNALILSDNKLSGS
                IPSQLSRLDRLRRLSLAGNDLSGTIPSELARFGGDDFSGNNGLCGKPLSRCGALNGRN
                LSIIIVAGVLGAVGSLCVGLVIFWWFFIREGSRKKKGYGAGKSKDDSDWIGLLRSHKL
                VQVTLFQKPIVKIKLGDLMAATNNFSSGNIDVSSRTGVSYKADLPDGSALAVKRLSAC
                GFGEKQFRSEMNKLGELRHPNLVPLLGYCVVEDERLLVYKHMVNGTLFSQLHNGGLCD
                AVLDWPTRRAIGVGAAKGLAWLHHGCQPPYLHQFISSNVILLDDDFDARITDYGLAKL
                VGSRDSNDSSFNNGDLGELGYVAPEYSSTMVASLKGDVYGFGIVLLELVTGQKPLSVI
                NGVEGFKGSLVDWVSQYLGTGRSKDAIDRSICDKGHDEEILQFLKIACSCVVSRPKER
                PTMIQVYESLKNMADKHGVSEHYDEFPLVFNKQEA"
CDS             complement(91271. .92998)
                /note="unknown, similar to ESTs emb|F14134, gb|T21189,
                gb|T43096, gb|T1189, gb|Z37659"
                /codon_start=1
                /evidence=not_experimental
                /product="T7N9.21"
                /db_xref="PID:g2213601"
                /translation="MTEYENGKKRKVRNKQQVKVQFLSQRYLILCFCCFFVLLFFLSS
                DRISTLSVRSDSLRPSLRVPTLSVLSSSMDSFHRGRFPPLSVEDRVQFPDHLLLILSH
                GIGKGEKNLVCVYRGVKEETLVLPSISSDEFDEFRSIVRCPNAPLNYSSSVDLQFRGD
                LVKKKMKKQSRRVHNWEKVGYEAVIDGDTVVVFVKGLTRRPHKESDPSYYKCQFEIEN
                SEEKEVTQAIAAAQEVVRCGLPESLKLNPEMMFRVSVIHIDPRGRTTPALPSVARIYG
                SDSIEKKEKKSGVKHELCVCTMLWNQAPFLREWIMYHSWLGVERWFIYDNNSDDGIQE
                EIELLSSENYNVSRHVWPWIKTQEAGFSHCAVRAKEECNWVGFFDVDEFYYFPTHRSQ
                GLPSKNALKSLVSNYTSWDLVGEIRTDCHSYGPSGLTSVPSQGVTVGYTCRQANPERH
                KSIIRPELLTSSLLNEVHHFQLKEGVGHMSLVESVAVVNHYKYQVWDTFKAKFYRRVA
                TYVVDWQENQNQGSKDRAPGLGTEAIEPPDWKRRFCEVWDTGLKDLVMSNFADQVTGY
                LPWQRQQQE"
CDS             complement(join(97226. .97606,97721. .98247,98334. .98433,
                98516. .99376))
                /note="unknown; similar to EST gb|N96069"
                /codon_start=1
                /evidence=not_experimental
                /product="T7N9.22"
                /db_xref="PID:g2213602"
                /translation="MSVHGRFPASPPISLSPSSSSTSPSSQSPSTPPDLKQRVIACLN
                KLADRDTLALASAELDSIARNLTHDSFSPFLNCIHNTDSSVKSPVRKQCVALLSVLSR
                YHGDSLTPHLAKMVSTVIRRLRDPDSSVRSACAVATADMSAHVTRQPFASVAKPLIET
                LIQEGDSNLQIGAALCLAASVDAATDPESEQLRKSLPKIGKLLKSDGFKAKAALLSAV
                GSIITAGGAGTKPVLDWLVPVLIEFLSSEDWAARKSAAEALGKVATAEDLASQYKKTC
                TTALESRRFDKVKSVRETMNRALNLWKEVSTDDEASLSPSRSSTDDGNIGCFSSVTRS
                STIDVGLKSARPKKVTPIMKRSPSLPVNRSYAATRQKENLPKRNQGNMTMLVEEASSV
                DNKGPHFTPVKKSSEETEEKANSGGPDIIKHTISEKSREDSKVSSFGGLRSGSRVAPC
                SDDGDSVKNCKDDVEESKKDSEELSLIREQLALIENQQSSLLDLLQKFMGTSQSGIQS
                LESRVSGLEMALDEISCDLAVSNGRVPRNSSGCAGDSCSKLPGTEFLSPKFWRKTEER
                PRNRNTANEMAAYDQGMRESTDTNNGQRGGSVFQKRSRRDQFQDCMHTTLQKPTTR"
CDS             join(105265. .105404,105539. .105663,106614. .106903)
                /note="unknown"
                /codon_start=1
                /evidence=not_experimental
                /product="T7N9.23"
                /db_xref="PID:g2213603"
                /translation="MAGEGSKPKATVDDAYAYIRTVKSTFHNDPDKYDDFMAIMKNFK
                ARKIDRNTCIEEVKELLKGHRDLISGFNAFLPKCLEIADWYNIEVLEAELQALLMAMQ
                HTWSKGYRKIIFEGDNRTVESLLNGNTRNFELHNLIIEIQHWRKKFSNAIIKWSFRST
                NKAADRLAKGELENNVTFVSHSTR"
CDS             join(107696. .107835,107952. .108198,108274. .108375,
                108464. .108535)
                /note="unknown"
                /codon_start=1
                /evidence=not_experimental
                /product="T7N9.24"
                /db_xref="PID:g2213604"
                /translation="MVERRVQVEPTLSDAHSYITAVKEAFHDEPTKYEEFIKLMNDIR
                DHGVDKASGIAKLTELIKGHPRLLRGLSFFFPQVNRDIHHEAKRTIILKDKATIPPEA
                AYRGAKSTYTKIKQIEPDWENFMNMLKTRFRSLDTHVVESFLKIMIMYDEGKKSEKEV
                QEEVVDLLYYHEDLIDKFFRLFNMRK"
CDS             join(110027. .110187,110323. .110453,110972. .111093)
                /note="unknown"
                /codon_start=1
                /evidence=not_experimental
                /product="T7N9.25"
                /db_xref="PID:g2213605"
                /translation="MVGGGSKHVGEGSKPRATKDDAYAYLRAVRAKFHNDSKKYDDFV
                AVMTNFKARRIDRDGCIKEVEQLLKGHRDLISGFNAFLPKCLEIKNYYFGAGEDLEVV
                YTAKRKLLARSVTKRSSSSSSSSVYFLFRSCEII"
BASE COUNT    35883 a  19310 c  20289 g  35908 t
ORIGIN


  Query Match             3.2%;  Score 41;  DB 38;  Length 111390;
  Best Local Similarity   49.5%;  Pred. No. 3.9;
  Matches  136;  Conservative    0;  Mismatches  135;  Indels    4;  Gaps    1;

Qy     662 ACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAAG 721
           ||| | |  | | |||| |    | ||   |  | ||  || || || | |   ||| ||
Db 109688 ACAAACCAGAAAATTTACTATGGGTATAAGGTTCACCACTAGAACTTGAATGATTTTAAT 109629

Qy     722 ACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAAT----CCA 777
             |  |||   |     |  ||  | | || ||||| |||     || ||||     ||
Db 109628 TTGAGCTGTATAGGTCTTTATATATATATAAAAAAAAATAGACCTTTTGCAATTGTATCA 109569

Qy     778 TAACATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATT 837
           ||| |   |    | | ||||| |||  ||  |   |||| |||  |    |  || | |
Db 109568 TAATACTAAATATAATAAACACATGTTAGTTGTAAATAAAAATTTCACATTAATTACAAT 109509

Qy     838 TGTTAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTAC 897
           |   | || | | |||| | |   | |  | |  ||   ||| || ||  |||||
Db 109508 TTACAGATTTTTATGTAAATTTTAATTATACATATATTTATGTAAATTTAGTAATAAATA 109449

Qy     898 ATGGAGACAGGTCAACCAAAGAGGGAGCTAGGCAA 932
           ||  | |  |   | |   | | | | ||  | ||
Db 109448 ATTTAAATTGTAAAGCTTTACAAGCATCTGTGAAA 109414


Search completed: May  3, 1999, 22:21:25
Job time: 2813 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 21:33:44 ; Search time 46.99 Seconds
(without alignments)
5088.631 Million cell updates/sec

Title: US-09-037-460-1
Perfect score: 1271
Sequence: 1 CTGCTTCCCACCAGCAAAGA.....GTTNCCTNTNTCTNAAGGAC 1271

Scoring table: IDENTITY_NUC

Searched: 240622 seqs, 94065609 residues

Database : N_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1263.2	99.4	1271	1	T34991	Vascular IBP-like
c	2	37.6	3.0	1198	1	T47647	Pineal gland speci
c	3	36.4	2.9	654	1	T47675	Human G protein ga
	4	36.4	2.9	2267	1	T94700	Murine Fisp12 cDNA
	5	36.4	2.9	2330	1	Q26422	Gene for beta-IG-M
	6	34.4	2.7	230	1	Q36033	Chicken nov gene e
	7	33.2	2.6	83	1	Q66926	PolydA detection p
	8	35.8	2.8	1975	1	Q36031	Chicken nov coding
	9	34.6	2.7	833	1	T51381	D. immitis VA5 nuc
c	10	34.6	2.7	833	1	T51376	D. immitis venom a
	11	35.2	2.8	2075	1	T04226	Connective tissue
	12	35.2	2.8	2075	1	T51234	Connective tissue
	13	35.2	2.8	2075	1	T45360	Human connective t
	14	35.2	2.8	2075	1	V38085	Human connective t
c	15	34	2.7	581	1	T69172	Trypanosoma cruzi
	16	34.6	2.7	1592	1	Q14501	Usp45 gene. Lactoc
	17	34.6	2.7	1920	1	Q13830	MSP gene in pUCRS
	18	35.2	2.8	4212	1	V65380	Connective tissue
	19	35.2	2.8	4214	1	T58534	Human connective t
	20	35.2	2.8	4214	1	T59618	Connective tissue
c	21	34.8	2.7	3238	1	Q83532	Sequence encoding
	22	33.8	2.7	1285	1	V29260	Human connective t
c	23	34.4	2.7	2769	1	T59196	Neuronal alpha-bun
	24	33.4	2.6	1042	1	V30303	Bacillus thuringie
	25	33.4	2.6	1045	1	V30301	Bacillus thuringie
	26	34	2.7	2418	1	Q27886	P.falciparum GBP13
	27	33.8	2.7	2105	1	T88022	Murine TRAF5 cDNA.
	28	32.8	2.6	684	1	Q36032	Chicken nov gene f
c	29	32.8	2.6	713	1	Q48773	Aphrodisine. DNA a
c	30	32.8	2.6	713	1	Q49210	Hamster Aphrodisin
	31	33.2	2.6	1723	1	T91745	Tobacco calcium/ca
c	32	34.8	2.7	16550	1	V21211	Methanococcus jann
c	33	32.8	2.6	1847	1	V59649	Human secreted pro
c	34	30.2	2.4	102	1	T26575	Human gene signatu
	35	32.8	2.6	2414	1	T04895	Rat SCF genomic DN
	36	31.4	2.5	552	1	N60612	Chloroplast promot
	37	31.8	2.5	1046	1	V30312	Bacillus thuringie
	38	33.2	2.6	5852	1	Q11710	Dictyostelium plas
c	39	31	2.4	428	1	Q60816	Human brain Expres
	40	32.2	2.5	1892	1	Q74063	The mouse beta-act
	41	32.8	2.6	4590	1	N60472	Sequence encoding
c	42	33	2.6	5852	1	Q11710	Dictyostelium plas
c	43	32	2.5	1835	1	T62654	DNA encoding prote
	44	32.8	2.6	4854	1	Q11538	Rat Stem Cell Fact
	45	31.6	2.5	1133	1	T74986	Human HLA DR alpha

ALIGNMENTS

```
RESULT   1
T34991
ID   T34991 standard; cDNA; 1271 BP.
AC   T34991;
DT   06-NOV-1996  (first entry)
DE   Vascular IBP-like growth factor cDNA.
KW   Vascular IBF-like growth factor; VIGF;
KW   insulin-like growth factor binding protein; agonist; antagonist;
KW   muscle wastage; osteoporosis; implant fixation; wound healing;
KW   therapy; diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             58. .612
FT                   /*tag= a
FT   signal_peptide  58. .120
FT                   /*tag= b
FT   mat_peptide     121. .609
FT                   /*tag= c
FT   misc_difference 653
FT                   /*tag= d
FT                   /note= "base 653 is given as y in the
FT                   specification"
FT   misc_difference 1096
FT                   /*tag= e
FT                   /note= "base 1096 is given as r in the
FT                   specification"
FT   misc_difference 1212
FT                   /*tag= f
FT                   /note= "base 1212 is given as n in the
FT                   specification"
FT   misc_difference 1225
FT                   /*tag= g
FT                   /note= "base 1225 is given as y in the
FT                   specification"
FT   misc_difference 1255
FT                   /*tag= h
FT                   /note= "base 1255 is given as n in the
FT                   specification"
FT   misc_difference 1259
FT                   /*tag= i
FT                   /note= "base 1259 is given as n in the
FT                   specification"
FT   misc_difference 1261
FT                   /*tag= j
FT                   /note= "base 1261 is given as n in the
FT                   specification"
FT   misc_difference 1265
FT                   /*tag= k
FT                   /note= "base 1265 is given as n in the
FT                   specification"
PN   WO9617931-A1.
PD   13-JUN-1996.
PF   09-DEC-1994; U14388.
PR   09-DEC-1994; WO-U14388.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Hastings GA, Rosen CA;
DR   WPI; 96-287176/29.
DR   P-PSDB; R98994.
PT   Human vascular insulin-like growth factor binding protein-like
PT   growth factor, and its nucleic acid sequence and (ant)agonists -
PT   used, e.g. to treat muscle wasting diseases or aid implant fixation,
PT   or limit excess connective tissue prodn. during wound healing.
PS   Claim 1; Page 42-43; 61pp; English.
CC   A cDNA clone (T34991) codes for human vascular insulin-like
CC   growth factor binding protein-like growth factor (R98994),
CC   or VIGF, a protein structurally related to the IBP and CCN
```

```
CC   families.  It was discovered in a cDNA library derived from human
CC   umbilical vein endothelial cells.  The cDNA can be used for
CC   prodn. of recombinant VIGF, e.g. by expression in E. coli, CHO
CC   or insect host cells.  It is also useful therapeutically e.g.
CC   to treat muscle wasting diseases or osteoporosis, or to design
CC   probes for the detection of diseases associated with under- or
CC   over-expression of VIGF.
SQ   Sequence   1271 BP;    360 A;    246 C;    337 G;    320 T;


  Query Match            99.4%;  Score 1263.2;  DB 1;  Length 1271;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1270;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60

Qy     61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120

Qy    121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180

Qy    181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240

Qy    241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300

Qy    301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360

Qy    361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420

Qy    421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480

Qy    481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540

Qy    541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600

Qy    601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAA 660

Qy    661 CACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAA 720

Qy    721 GACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780

Qy    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840

Qy    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900

Qy    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960

Qy    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020

Qy   1021 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGATGGGGGGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080

Qy   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140

Qy   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200

Qy   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260

Qy   1261 NTCTNAAGGAC 1271
          ||||||||||| 
Db   1261 NTCTNAAGGAC 1271


RESULT   2
T47647/c
ID   T47647 standard; cDNA; 1198 BP.
AC   T47647;
DT   17-SEP-1997  (first entry)
DE   Pineal gland specific gene-1 encoding cDNA.
KW   PGSG-1; pineal gland; epiphysis cerebri; tumour; precocious puberty;
KW   hydrocephalus; papilledema; intracranial pressure; circadian rhythm;
KW   pituitary secretion; luteinising hormone; growth hormone;
KW   follicular stimulating hormone; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  37. .99
FT                   /*tag= a
FT                   /note= "Putative leader sequence"
FT   mat_peptide     100. .1071
FT                   /*tag= b
FT                   /product= Pineal_gland_specific_gene-1
PN   WO9639158-A1.
PD   12-DEC-1996.
PF   05-JUN-1995; U07067.
PR   05-JUN-1995; WO-U07067.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   He WW,  Rosen CA;
DR   WPI; 97-042840/04.
DR   P-PSDB; W09405.
PT   Pineal gland specific gene-1 and corresponding protein - used in the
PT   treatment of pineal tumours and alleviation of side effect, e.g.
PT   precocious puberty, hydrocephalus etc.
PS   Claim 1; Page 42; 56pp; English.
CC   The present sequence represents a novel isolated polynucleotide
CC   pineal gland specific gene-1 (PGSG-1), which was derived from a human
CC   pineal gland tissue cDNA library. The PGSG-1 polypeptide may be used
CC   to treat pineal tumours and thereby treat the side effects, including
CC   precocious puberty, hydrocephalus, papilledema and other signs of
CC   increased intracranial pressure. PGSG-1 and its protein may be used to
CC   regulate biological rhythms, in particular circadian rhythms, and to
CC   regulate pituitary secretion of hormones which regulate the onset of
CC   puberty, e.g. luteinising hormone, follicular stimulating hormone (FSH)
CC   and growth hormone (GH) released by the pituitary. The (ant)agonists
CC   which act against the protein may also be used to regulate the secretion
CC   of these hormones. The PGSG-1 gene and proteins may also be used to
CC   diagnose a mutation in the PGSG-1 gene and hence susceptibility to a
```

```
CC   disease mentioned above.
SQ   Sequence   1198 BP;    300 A;    320 C;    279 G;    299 T;


  Query Match              3.0%;  Score 37.6;  DB 1;  Length 1198;
  Best Local Similarity   76.7%;  Pred. No. 0.28;
  Matches   46;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy   1100 TTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAAACAACCAGAAAACCCCTGA 1159
           |||||||||||||||| | |||| || |||  |||||| | || | ||| |||| | | |
Db   1194 TTTTTTTTTTTTTTTTTGTCTTTTGGCAAAAGCAAAATCAGACTAACAGGAAACACATCA 1135


RESULT   3
T47675/c
ID   T47675 standard; cDNA; 654 BP.
AC   T47675;
DT   19-MAY-1997  (first entry)
DE   Human G protein gamma-11 subunit cDNA.
KW   G protein; signal transduction; agonist; antagonist; diagnosis;
KW   therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   5'utr           1. .107
FT                   /*tag= a
FT   cds             108. .329
FT                   /*tag= b
FT   3'utr           330. .654
FT                   /*tag= c
FT   polya_signal    607. .612
FT                   /*tag= d
PN   WO9637513-A1.
PD   28-NOV-1996.
PF   22-MAY-1995; U06406.
PR   22-MAY-1995; WO-U06406.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (WEIS-) WEIS CENT RES.
PI   Kunsch CA,  Robishaw JD;
DR   WPI; 97-021140/02.
DR   P-PSDB; W09418.
PT   New nucleic acid encoding gamma subunits of human G protein - used
PT   to detect mutation(s) in, or altered levels of, the subunits and to
PT   screen for agonists and antagonists of G protein interaction
PS   Claim 1; Page 29; 40pp; English.
CC   cDNA clones (T47669-75) respectively code for human G protein
CC   gamma-2, gamma-3, gamma-4, gamma-5, gamma-7, gamma-10 and gamma-11
CC   subunits (W09412-18).  Several human cDNA libraries were partially
CC   sequenced to identify expressed sequence tags (EST).  By matching
CC   sequences of EST to genes of known structure, the 7 new subunit
CC   sequences were isolated.  The cDNA clone for gamma-11 has been
CC   deposited as ATCC 97139.  The isolated cDNA clones can be used to
CC   produce recombinant gamma subunits in prokaryotic or eukaryotic
CC   host cells for use in the development of therapeutic and diagnostic
CC   agents.  The nucleic acids can also be used to produce probes to
CC   detect mutations in human G protein subunit sequences and for
CC   chromosome identification.
SQ   Sequence   654 BP;    231 A;    116 C;    142 G;    165 T;


  Query Match              2.9%;  Score 36.4;  DB 1;  Length 654;
  Best Local Similarity   58.2%;  Pred. No. 0.45;
  Matches   64;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy   1089 TTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAAACAACCAG 1148
           |||  || |  |||||||||||||||||    |  |  |||| |  |  | || || | ||
Db    654 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAGACAAAACTTTATTTGAAGCATCAAG 595

Qy   1149 AAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAA 1198
            ||    ||||| |||||   ||| |  |    | |    ||||   |||
Db    594 CAATATACTGAAAGAAGTGTCATGCTCAAGTAATGTTCTTATTTCCATAA 545


RESULT   4
T94700
ID   T94700 standard; cDNA; 2267 BP.
AC   T94700;
DT   27-MAR-1998  (first entry)
DE   Murine Fispl2 cDNA.
KW   Fispl2; cysteine rich protein; mouse; Cyr61;
KW   extracellular matrix signalling molecule; cell adhesion;
KW   cell migration; cell proliferation; angiogenesis; chrondrogenesis;
KW   oncogenesis; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             138. .1284
FT                   /*tag= a
PN   WO9733995-A2.
PD   18-SEP-1997.
PF   14-MAR-1997; U04193.
PR   15-MAR-1996; US-013958.
PA   (MUNI-) MUNIN CORP.
PI   Lau LF;
DR   WPI; 97-470875/43.
DR   P-PSDB; W35731.
PT   Isolated and purified cysteine rich protein 61, Cyr61 - useful to
PT   modulate e.g. haematostasis, induce wound healing, promote organ
PT   regeneration etc
PS   Example 2; Page 113-114; 133pp; English.
CC   This cDNA clone codes for murine Fispl2 (see W35731), an
CC   extracellular matrix signalling molecule (ECM) that exhibits
CC   structural similarity to Cyr61 (see W35730) and which, like
CC   Cyr61, influences cell adhesion, proliferation and migration.
CC   The human orthologue of Fispl2 is connective tissue growth
CC   factor.  Fispl2 polynucleotides can be used for the production
CC   of Fispl2 polypeptides by recombinant methods.  Polypeptide
CC   compositions are provided that comprise mammalian ECM signalling
CC   molecules, peptide fragments, inhibitory peptides capable of
CC   interacting with receptors for ECM signalling molecules, and
CC   antibody products.  Further provided are methods for using
CC   mammalian ECM signalling molecules to screen for, and/or modulate
CC   disorders associated with angiogenesis, chondrogenesis and
CC   oncogenesis; ex vivo methods for using ECM signalling molecules
CC   to prepare blood products are also provided.
SQ   Sequence   2267 BP;    570 A;    568 C;    582 G;    547 T;


  Query Match              2.9%;  Score 36.4;  DB 1;  Length 2267;
  Best Local Similarity   66.7%;  Pred. No. 0.89;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy    174 CAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGC 233
          | ||||   | ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    260 CTGCCCCGCCGGCGTGAGCCTGGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCAA 319

Qy    234 AGGGCGGGGAGAAACTTG 251
            || |||||||   ||
Db    320 GCAGCTGGGAGAACTGTG 337


RESULT   5
Q26422
ID   Q26422 standard; DNA; 2330 BP.
AC   Q26422;
DT   18-JAN-1993  (first entry)
DE   Gene for beta-IG-M2.
KW   Transforming growth factor beta; induced; CEF-10; v-src; chicken;
KW   embryo; fibroblasts; TGF-beta; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             204. .1247
FT                   /*tag= a
PN   EP-495674-A.
PD   22-JUL-1992.
PF   17-JAN-1992; 300429.
PR   18-JAN-1991; US-642991.
```

```
PR   10-JAN-1992; US-816270.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Brunner AM, Chinn J, Neubauer MG, Purchio AF;
DR   WPI; 92-243508/30.
DR   P-PSDB; R25565.
PT   TGF-beta induced gene family - encodes proteins involved in
PT   growth and differentiation effects of TGF-beta-1
PS   Claim 7; Fig 2; 35pp; English.
CC   The DNA encoding mouse beta-IG-M2 was obtd. from AKR-2B mouse cells
CC   induced with TGF-beta1 and cyclohexamide.  Poly RNA extracted from
CC   these cells was used  to create a cDNA library which was screened
CC   using two probes.  The probes were prepd. from untreated AKR-2B mRNA
CC   and AKR-2B mRNA from cells treated with cyclohexamide and TGF-beta1.
CC   Hybridising colonies were isolated and two clones (beta-IG-M1 and
CC   beta-IG-M2) were then sequenced.  The DNA encodes proteins that
CC   have a 80 and 50 percent homology respectively with the CEF-10
CC   protein induced by v-src in chicken embryo fibroblasts.  The
CC   proteins encoded by TGF-beta induced genes are likely to be
CC   involved in mediation of the biological effects of TGF-beta
CC   relating to cell growth and differentiation.
CC   See also Q26421.
SQ   Sequence   2330 BP;    589 A;    598 C;    595 G;    548 T;


  Query Match             2.9%;  Score 36.4;  DB 1;  Length 2330;
  Best Local Similarity   66.7%;  Pred. No. 0.9;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy    174 CAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGC 233
          | ||||   | ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    326 CTGCCCCGCCGGCGTGAGCCTGGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCAA 385

Qy    234 AGGGCGGGGAGAAACTTG 251
            || |||||||   ||
Db    386 GCAGCTGGGAGAACTGTG 403


RESULT   6
Q36033
ID   Q36033 standard; cDNA; 230 BP.
AC   Q36033;
DT   24-MAY-1993  (first entry)
DE   Chicken nov gene exon 2 fragment III.
KW   avian nephroblastoma; avian myeloblastoma virus;
KW   stringent hybridisation; ss.
OS   Gallus domesticus.
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
DR   P-PSDB; R31600.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 2; Page 27; 67pp; French.
CC   The chicken nov gene is stimulated in avian nephroblastoma induced
CC   by avian myeloblastoma virus but not in normal adult kidney. A
CC   1975bp cDNA sequence was isolated from a gene bank prepared from
CC   chicken embryonic fibroblasts screened with a tumour-derived probe.
CC   Fragment III is derived from the 2nd. exon of the nov gene;
CC   nucleotide sequences which hybridise to Fragment III under stringent
CC   conditions (i.e. 50% formamide, 5 x SCC) are claimed. The claimed
CC   sequences preferably encode a protein with at least 70% homology to
CC   R31600 which is encoded by Fragment III.
SQ   Sequence   230 BP;    18 A;    89 C;    97 G;    26 T;


  Query Match             2.7%;  Score 34.4;  DB 1;  Length 230;
  Best Local Similarity   63.1%;  Pred. No. 0.96;
  Matches   53;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy    173 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 232
          || || | |   |   |    | ||||| |||| ||| ||||||||||  ||||||||
Db     67 GCTGCGCCCCGGGAGTGCCCGCCGTGCTGGACGGCTGCGGCTGCTGCCTGGTGTGCGCCC 126

Qy    233 CAGGGCGGGGAGAAACTTGCTACC 256
              ||| || || |  |||| ||
Db    127 GGCAGCGCGGCGAGAGCTGCTCCC 150


RESULT   7
Q66926
ID   Q66926 standard; DNA; 83 BP.
AC   Q66926;
DT   25-JAN-1995  (first entry)
DE   PolydA detection probe containing flourophores.
KW   Target sequence; binding ability; hairpin forming; probes;
KW   imperfect hairpin; acceptor label moiety; donor moiety label;
KW   fluoresence; fluorophores; specificity; base pair mismatches;
KW   competitive arms; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   modified_base   1
FT                   /*tag=  a
FT                   /note= "Three-site aminoalkyl-derivatised
FT                   oligonucleotide with a 6-carbon linker arm
FT                   attached to an additional fluorophore molecule"
FT   stem_loop       1. .41
FT                   /*tag=  b
FT                   /note= "Forms imperfect hairpin with the competitive
FT                   arm containing 4 mismatches"
FT   misc_difference 42
FT                   /*tag=  c
FT                   /note= "6-carbon molecular linker containing a
FT                   flourescein molecule and occupying the same
FT                   space as a base"
FT   stem_loop       43. .83
FT                   /*tag=  d
FT                   /note= "Forms imperfect hairpin with the competitive
FT                   arm containing 4 mismatches"
FT   modified_base   83
FT                   /*tag=  e
FT                   /note= "Three-site aminoalkyl-derivatised
FT                   oligonucleotide with a 6-carbon linker arm
FT                   attached to an additional fluorophore molecule"
PN   EP-601889-A.
PD   15-JUN-1994.
PF   10-DEC-1993; 310007.
PR   10-DEC-1992; US-990298.
PA   (MAIN-) MAINE MEDICAL CENT RES INST.
PI   Bagwell BC;
DR   WPI; 94-185245/23.
PT   Nucleic acid probe for use in DNA or RNA hybridisation assays -
PT   comprises a nucleotide sequence which is capable of forming one
PT   or more hairpins
PS   Disclosure; Fig 4; 25pp; English.
CC   This sequence represents a hairpin forming probe of the invention.  The
CC   probes comprise a segment complementary to the target nucleotide and are
CC   capable of forming at least 1 imperfect hairpin.  The probes contain at
CC   least one acceptor label moiety and at least one donor moiety label
CC   which are covalently attached to the nucleotide sequence so that when
CC   the hairpins are formed the moieties are in close proximity to allow
CC   resonance energy transfer between them.  This causes a reduction in the
CC   fluoresence of the two fluorophores.  The target sequence contains at
CC   least 22 nucleotides to ensure specificity and avidity of the probe-
CC   target hybridisation.  The hairpins formed are imperfect hairpins and
CC   the intentional base pair mismatches are introduced into the competitive
CC   arms such that the specificity sequence will favour binding to the
CC   target sequence.  Therefore when the probe interacts with the target
CC   sequence the competitive arm is displaced, increasing the distance
CC   between the fluorophores, resulting in a change in fluorescent emission.
SQ   Sequence   83 BP;    26 A;    4 C;    9 G;    43 T;
```

```
  Query Match             2.6%;  Score 33.2;  DB 1;  Length 83;
  Best Local Similarity   74.5%;  Pred. No. 1.2;
  Matches   41;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy   1089 TTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAAACA 1143
          |||  || |  || ||||||||||||||    |||||  |||||||||| |||| |
Db     29 TTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTGGTGAAAGTCAAAACAAAAAA 83


RESULT   8
Q36031
ID   Q36031 standard; cDNA; 1975 BP.
AC   Q36031;
DT   24-MAY-1993  (first entry)
DE   Chicken nov coding sequence.
KW   avian nephroblastoma; avian myeloblastoma virus; ss.
OS   Gallus domesticus.
FH   Key             Location/Qualifiers
FT   cds             24. .1079
FT                   /*tag=  a
FT   polya_signal    1814. .1818
FT                   /*tag=  b
FT   polya_signal    1932. .1937
FT                   /*tag=  c
FT   signal_peptide  24. .95
FT                   /*tag=  d
FT   exon            1. .92
FT                   /*tag=  e
FT                   /number= 1
FT   exon            93. .321
FT                   /*tag=  f
FT                   /number= 2
FT   exon            322. .573
FT                   /*tag=  g
FT                   /number= 3
FT   exon            574. .782
FT                   /*tag=  h
FT                   /number= 4
FT   exon            783. .1975
FT                   /*tag=  i
FT                   /number= 5
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
DR   P-PSDB; R31599.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 1; Fig 1; 67pp; French.
CC   The chicken nov gene is stimulated in avian nephroblastoma induced
CC   by avian myeloblastoma virus but not in normal adult kidney. The
CC   1975bp cDNA sequence was isolated from a gene bank prepared from
CC   chicken embryonic fibroblasts screened with a tumour-derived probe.
CC   The cDNA was cloned in Bluescript KS+ to form plasmid pC1K which was
CC   labelled and used to screen a human placental DNA library. A nov-
CC   analogous gene was isolated from the human library. Nucleotide
CC   sequences which hybridise to the chicken nov gene under stringent
CC   conditions (i.e. 50% formamide, 5 x SCC) are claimed.
SQ   Sequence   1975 BP;    522 A;    451 C;    486 G;    516 T;


  Query Match             2.8%;  Score 35.8;  DB 1;  Length 1975;
  Best Local Similarity   52.9%;  Pred. No. 1.2;
  Matches  101;  Conservative    0;  Mismatches   87;  Indels    3;  Gaps    1;

Qy     66 CGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAATAA 125
          |||| ||||||||  |  |||||||| | |   |   |  ||||  ||   ||
Db     53 CGTCCTGCTGCTGCTCCTGCTCCTCCTCCGGCCGTGCGAGGTGAGCGGGCGGGAGGCGGC 112

Qy    126 TTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCGCTG 185
          |   |   |  ||||       | |||   |  |   |||       || || | |   |
Db    113 GTGCCCCCGGCCCTGCGGCGGGCGCTG---CCCCGCGGAGCCGCCGCGCTGCGCCCCGGG 169

Qy    186 CAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGGAGA 245
          |    | ||||| |||| ||| ||||||||||  ||||||||      ||| || ||
Db    170 AGTGCCCGCCGTGCTGGACGGCTGCGGCTGCTGCCTGGTGTGCGCCCGGCAGCGCGGCGA 229

Qy    246 AACTTGCTACC 256
          |  |||| ||
Db    230 GAGCTGCTCCC 240


RESULT   9
T51381
ID   T51381 standard; cDNA; 833 BP.
AC   T51381;
DT   22-APR-1997  (first entry)
DE   D. immitis VA5 nucleic acid nDiVA833 complementary sequence.
KW   Venom allergen antigen 5-like gene; VA5; helminth; parasite;
KW   nDiVA833; PDiVA221; vaccine; heartworm; ss.
OS   Dirofilaria immitis.
PN   WO9637218-A1.
PD   28-NOV-1996.
PF   23-MAY-1996; U07709.
PR   23-MAY-1995; US-450944.
PA   (HESK-) HESKA CORP.
PI   Tripp CA,  Wisnewski N;
DR   WPI; 97-020935/02.
PT   Nucleic acid encoding helminth venom allergen antigen 5-like protein
PT   - pref. from Dirofilaria immitis or Onchocerca volvulus, useful esp.
PT   in vaccines to prevent helminth infection
PS   Claim 12; Page 95-96; 124pp; English.
CC   A cDNA sequence (T51381) is the complementary strand of the
CC   Dirofilaria immitis nucleic acid nDiVA833 (T51376), which codes
CC   for venom allergen antigen 5-like protein VA5 (W11479).  Parasitic
CC   helminth VA5-like gene nucleic acids (T51376-97) can be used to
CC   produce VA5-like proteins (W11480-86) either in vitro or in vivo.
CC   The proteins elicit an immune response that protects an animal from
CC   helminth infection.
SQ   Sequence   833 BP;    259 A;    165 C;    110 G;    299 T;


  Query Match             2.7%;  Score 34.6;  DB 1;  Length 833;
  Best Local Similarity   52.1%;  Pred. No. 1.7;
  Matches   73;  Conservative    1;  Mismatches   66;  Indels    0;  Gaps    0;

Qy   1100 TTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAAACAACCAGAAAACCCCTGA 1159
          |||||||||||||||| |   |||   |   |  ||  |  |  |  |  |  |||| ||
Db      5 TTTTTTTTTTTTTTTTGCAGTTGTTTATTTTATAATAATTAATAAAATATTACCCTTGT 64

Qy   1160 AGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACAAACAGCTTTGANCTGAGAG 1219
          |  |  | | || | ||||   |  |||| |   ||  || |  |||   |   |
Db     65 AAAATATCAAATATATGAACAAAAATTAAATATTGAAAATGAAAATTTTTTCATTTTTTG 124

Qy   1220 CAATTYCAAAAGGCTGCTGA 1239
          ||   :| | |  ||||||
Db    125 CACAATCCAGATTTTGCTGA 144


RESULT  10
T51376/c
ID   T51376 standard; cDNA; 833 BP.
AC   T51376;
DT   22-APR-1997  (first entry)
DE   D. immitis venom allergen antigen 5-like gene clone nDiVA833.
KW   Venom allergen antigen 5-like gene; VA5; helminth; parasite;
KW   nDiVA833; PDiVA221; vaccine; heartworm; ss.
OS   Dirofilaria immitis.
FH   Key             Location/Qualifiers
FT   cds             54. .719
FT                   /*tag=  a
```

```
FT                   /note= "the coding sequence (minus the stop codon)
FT                   is represented by sequence nDiVA663"
FT   signal_peptide  54. .101
FT                   /*tag= b
FT   mat_peptide     102. .716
FT                   /*tag= c
FT                   /note= "the mature peptide-encoding sequence is
FT                   represented by sequence nDiVA205"
PN   WO9637218-A1.
PD   28-NOV-1996.
PF   23-MAY-1996; U07709.
PR   23-MAY-1995; US-450944.
PA   (HESK-) HESKA CORP.
PI   Tripp CA, Wisnewski N;
DR   WPI; 97-020935/02.
DR   P-PSDB; W11479.
PT   Nucleic acid encoding helminth venom allergen antigen 5-like protein
PT   - pref. from Dirofilaria immitis or Onchocerca volvulus, useful esp.
PT   in vaccines to prevent helminth infection
PS   Claim 4; Page 89-90; 124pp; English.
CC   A cDNA clone (T51376), designated nDiVA833, codes for the full-
CC   length venom allergen antigen 5-like protein VA5 (W11479),
CC   designated PDiVA221, of Dirofilaria immitis. It was isolated from
CC   a D. immitis cDNA library constructed from 3rd stage larvae by
CC   screening with immune dog serum. Parasitic helminth VA5-like gene
CC   nucleic acids (T51376-97) can be used to produce VA5-like proteins
CC   (W11480-86) either in vitro or in vivo. The proteins elicit an
CC   immune response that protects an animal from helminth infection.
SQ   Sequence  833 BP;  299 A;  110 C;  165 G;  259 T;


  Query Match              2.7%;  Score 34.6;  DB 1;  Length 833;
  Best Local Similarity   52.1%;  Pred. No. 1.7;
  Matches   73;  Conservative    1;  Mismatches   66;  Indels    0;  Gaps    0;

Qy   1100 TTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAAACAACCAGAAAACCCCTGA 1159
          |||||||||||||||| |   |||   |   |  ||  |  |  |  | |  |||| ||
Db    829 TTTTTTTTTTTTTTTTTGCAGTTGTTTATTTTATAATAATTAATAAAATATTACCCTTGT 770

Qy   1160 AGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACAAACAGCTTTGANCTGAGAG 1219
          |  |  | | || | ||||    |   |||| |    ||  || |  |||    |    |
Db    769 AAAATATCAAATATATGAACAAAAATTAAATATTGAAAATGAAAATTTTTTCATTTTTTG 710

Qy   1220 CAATTYCAAAAGGCTGCTGA 1239
          ||   :| | |   ||||||
Db    709 CACAATCCAGATTTTGCTGA 690


RESULT  11
T04226
ID   T04226 standard; cDNA; 2075 BP.
AC   T04226;
DT   12-JUN-1996  (first entry)
DE   Connective tissue growth factor cDNA.
KW   Connective tissue growth factor; CTGF; wound healing; vulnerary;
KW   cell proliferation; cancer; fibrosis; atherosclerosis; diagnosis
KW   therapy; mitogen; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             130. .1179
FT                   /*tag= a
PN   US5408040-A.
PD   18-APR-1995.
PF   30-AUG-1991; 752427.
PR   30-AUG-1991; US-752427.
PR   14-DEC-1993; US-167628.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM, Grotendorst GR;
DR   WPI; 95-161147/21.
DR   P-PSDB; R79964.
PT   New connective tissue growth factor - used to develop prods. for
PT   wound healing and for diagnosis and therapy of cell proliferative
PT   disorders.
PS   Example 5; Column 15-18; 12pp; English.
CC   A cDNA clone (T04226), designated DB60R32, codes for human
CC   connective tissue growth factor (CTGF) (R79964), a protein that
CC   has mitogenic and chemotactic activity for connective tissue
CC   cells and which binds to the platelet-derived growth factor
CC   (PDGR) receptor. The cDNA was obtd. from a library of human
CC   umbilical vein endothelial cell cDNA in vector lambda gt11 by
CC   screening with an anti-PDGF antibody. The cDNA is useful for
CC   prodn. of recombinant CTGF or in the design of antisense or
CC   ribozyme constructs used to treat disorders associated with
CC   overgrowth of tissue cells, such as cancer, fibrotic diseases
CC   and atherosclerosis.
SQ   Sequence  2075 BP;  491 A;  558 C;  546 G;  480 T;


  Query Match              2.8%;  Score 35.2;  DB 1;  Length 2075;
  Best Local Similarity   65.0%;  Pred. No. 1.9;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    173 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 232
          || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    254 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 313

Qy    233 CAGGGCGGGGAGAAACTTGC 252
          || ||| ||    |||
Db    314 AGCAGCTGGGCGAGCTGTGC 333


RESULT  12
T51234
ID   T51234 standard; cDNA; 2075 BP.
AC   T51234;
DT   18-MAR-1997  (first entry)
DE   Connective tissue growth factor coding sequence.
KW   Connective tissue growth factor; CTGF; human; connective tissue cell;
KW   proliferative disease; platelet-derived growth factor; PDGF; development;
KW   tissue growth; repair; umbilical vein endothelial cell; HUVE cell;
KW   antibody; wound healing; cancer; fibrotic disease; atherosclerosis;
KW   inhibitor; protease degradation; growth factor; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             130. .1179
FT                   /*tag= a
FT                   /product= connective tissue growth factor
PN   US5585270-A.
PD   17-DEC-1996.
PF   30-AUG-1991; 752427.
PR   30-AUG-1991; US-752427.
PR   14-DEC-1993; US-167628.
PR   10-FEB-1995; US-386680.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM, Grotendorst GR;
DR   WPI; 97-051180/05.
DR   P-PSDB; W11302.
PT   New nucleic acid encoding connective tissue growth factor - useful
PT   for accelerating wound healing, also for diagnosis and treatment of
PT   proliferative disease
PS   Claim 10; Column 15-18; 11pp; English.
CC   This sequence represents the cDNA encoding the human connective tissue
CC   growth factor (CTGF). CTGF is related immunologically and biologically
CC   to platelet-derived growth factor (PDGF), but is encoded by an unrelated
CC   gene. CTGF is thought to play a significant role in the normal
CC   development, growth, and repair of human tissue, similarly to PDGF. This
CC   sequence was isolated by screening a cDNA library from human umbilical
CC   vein endothelial (HUVE) cells with anti-PDGF antibodies. CTGF can be
CC   used to accelerate wound healing. Also, elevated levels of CTGF may be
CC   diagnostic of proliferative diseases involving outgrowth of connective
CC   tissue cells, such as cancer, fibrotic disease and atherosclerosis. All
CC   of these diseases can be treated with reagents reactive with CTGF, such
CC   as antibodies (which can also serve as assay reagents). Antisense
CC   nucleic acids, and ribozymes could also be used to inhibit CTGF
CC   production. The advantage with using CTGF is that it is more stable, and
CC   less susceptible to protease degradation than PDGF, and other growth
```

```
CC   factors involved in wound healing.  This is believed to be due to the
CC   high Cys content.
SQ   Sequence   2075 BP;    491 A;    558 C;    546 G;    480 T;


  Query Match              2.8%;  Score 35.2;  DB 1;  Length 2075;
  Best Local Similarity   65.0%;  Pred. No. 1.9;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    173 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 232
          || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    254 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 313

Qy    233 CAGGGCGGGGAGAAACTTGC 252
             || ||| ||    |||
Db    314 AGCAGCTGGGCGAGCTGTGC 333


RESULT  13
T45360
ID   T45360 standard; cDNA; 2075 BP.
AC   T45360;
DT   26-APR-1997  (first entry)
DE   Human connective tissue growth factor cDNA.
KW   Connective tissue growth factor; CTGF; mitogen; cell proliferation;
KW   wound healing; cancer; tumour; fibrosis; glaucoma; atherosclerosis;
KW   diagnosis; therapy; antisense; triple helix; ribozyme; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             130. .1179
FT                   /*tag= a
PN   WO9638172-A1.
PD   05-DEC-1996.
PF   31-MAY-1996; U08140.
PR   31-MAY-1996; WO-U08140.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM,  Grotendorst GR;
DR   WPI; 97-042659/04.
DR   P-PSDB; W09089.
PT   Connective tissue growth factor coding sequence and protein  - used
PT   in the treatment of proliferative disorders and to accelerate wound
PT   healing
PS   Claim 20; Page 50-52; 76pp; English.
CC   A cDNA clone (T45360) codes for novel human connective tissue growth
CC   factor (CTGF) (W09089), a PDGF-immunorelated protein that may play a
CC   significant role in the normal development, growth and repair of
CC   human tissue and which probably functions as a growth factor in
CC   wound healing.  CTGF may be involved in diseases in which there is
CC   an overgrowth of connective tissue cells, such as cancer, tumour
CC   formation and growth, fibrotic diseases (e.g. pulmonary fibrosis,
CC   kidney fibrosis, glaucoma) and atherosclerosis.  The cDNA clone was
CC   isolated from a HUVEC cDNA library using anti-PDGF antibody.  CTGF
CC   nucleic acids can be used for recombinant prodn. of CTGF and as
CC   probes to detect CTGF mRNA.  CTGF genomic DNA (T58534) has also
CC   been isolated.  A proliferative disorder may be treated using e.g.
CC   a CTGF antisense, ribozyme or triplex agent.
SQ   Sequence   2075 BP;    491 A;    558 C;    546 G;    480 T;


  Query Match              2.8%;  Score 35.2;  DB 1;  Length 2075;
  Best Local Similarity   65.0%;  Pred. No. 1.9;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    173 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 232
          || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    254 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 313

Qy    233 CAGGGCGGGGAGAAACTTGC 252
             || ||| ||    |||
Db    314 AGCAGCTGGGCGAGCTGTGC 333


RESULT  14
V38085
ID   V38085 standard; cDNA; 2075 BP.
AC   V38085;
DT   15-SEP-1998  (first entry)
DE   Human connective tissue growth factor encoding cDNA.
KW   Human; connective tissue growth factor; CTGF; PDGF; diagnosis; cancer;
KW   platelet derived growth factor; ameliorating cell proliferative disorder;
KW   atherosclerosis; fibrotic disease; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             130. .1179
FT                   /*tag= a
FT                   /product= "connective tissue growth factor"
PN   US5783187-A.
PD   21-JUL-1998.
PF   11-SEP-1996; 712302.
PR   30-AUG-1991; US-752427.
PR   14-DEC-1993; US-167628.
PR   11-SEP-1996; US-712302.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM, Grotendorst GR;
DR   WPI; 98-426958/36.
DR   P-PSDB; W62084.
PT   Ameliorating cell proliferative disorder associated with connective
PT   tissue growth factor - comprises the administration of an antibody
PT   which binds to connective tissue growth factor and not to
PT   platelet-derived growth factor
PS   Example 6; Column 15-18; 11pp; English.
CC   A method has been developed for ameliorating a cell proliferative
CC   disorder associated with connective tissue growth factor (CTGF). The
CC   method comprises the administration of an antibody or its fragment that
CC   binds to CTGF and not to platelet-derived growth factor (PDGF), to the
CC   site of the disorder. CTGF is related immunologically and biologically
CC   to PDGF. The present sequence encodes CTGF. The method is used to treat
CC   conditions involving the overgrowth of connective tissue cells such
CC   as cancer, atherosclerosis and other fibrotic diseases.
SQ   Sequence   2075 BP;    491 A;    558 C;    546 G;    480 T;


  Query Match              2.8%;  Score 35.2;  DB 1;  Length 2075;
  Best Local Similarity   65.0%;  Pred. No. 1.9;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    173 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 232
          || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    254 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 313

Qy    233 CAGGGCGGGGAGAAACTTGC 252
             || ||| ||    |||
Db    314 AGCAGCTGGGCGAGCTGTGC 333


RESULT  15
T69172/c
ID   T69172 standard; cDNA; 581 BP.
AC   T69172;
DT   12-JAN-1998  (first entry)
DE   Trypanosoma cruzi antigen (ubiquitin) cDNA.
KW   Antigen; epitope; vaccine; protective immunity; Chagas disease;
KW   diagnosis; therapy; immunoassay; ubiquitin; ss.
OS   Trypanosoma cruzi Tulahean strain C2.
FH   Key             Location/Qualifiers
FT   CDS             45. .431
FT                   /*tag= a
PN   WO9718475-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18624.
PR   14-NOV-1995; US-557309.
PA   (CORI-) CORIXA CORP.
PI   Houghton RL,  Lodes MJ,  Reed SG,  Skeiky YAW;
DR   WPI; 97-289413/26.
DR   P-PSDB; W26534.
PT   Diagnosing Trypanosoma cruzi infection by detecting antibodies to
```

PT novel antigens - which are useful in vaccines to provide protective
PT immunity against Chagas' disease
PS Claim 1; Page 50-51; 110pp; English.
CC This novel DNA molecule was isolated by screening a Trypanosoma
CC cruzi cDNA expression library with pools of sera from infected
CC individuals. It encodes T. cruzi ubtiquitin (see W26534). 22
CC Isolated genomic DNA of cDNA molecules (T69151-72) encode full-
CC length antigens (see W26530-41), or epitope-containing repeat
CC sequences (see W19094-102, W19079-86 and W26542-44) of native
CC antigens, that can be used in a variety of immunoassays for
CC detecting T. cruzi infection in a blood, serum, plasma, saliva,
CC cerebrospinal fluid or urine sample. The polypeptides are also
CC useful in vaccines and pharmaceutical compositions for inducing
CC protective immunity against Chagas disease. The isolated DNA
CC molecules can be used for recombinant production of the antigenic
CC polypeptides.
SQ Sequence 581 BP; 160 A; 139 C; 155 G; 127 T;

Query Match 2.7%; Score 34; DB 1; Length 581;
Best Local Similarity 63.4%; Pred. No. 2.1;
Matches 52; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy   1089 TTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAAACAACCAG 1148
          |||  || |  ||||||||||||||||    |||  |||||   || |  ||  ||| ||
Db    580 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAAAAAAAAACAACAACAAACAAG 521

Qy   1149 AAAACCCCTGAAGGAAGTAAGA 1170
          | | |  || ||  ||  || |
Db    520 ACATCTTCTCAATAAAAAAAAA 499
```

Search completed: May 3, 1999, 22:00:48
Job time: 1624 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 17:43:40 ; Search time 47.06 Seconds
(without alignments)
191.088 Million cell updates/sec

Title: US-09-037-460-2_COPY_1_163
Perfect score: 820
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....NRFVSLTEHDMASGDGNIVR 163

Scoring table: PAM150

Searched: 180763 seqs, 55169189 residues

Database :
SPTREMBL_8:*
1: sp_fungi:*
2: sp_human:*
3: sp_invertebrate:*
4: sp_mammal:*
5: sp_mhc:*
6: sp_organelle:*
7: sp_phage:*
8: sp_plant:*
9: sp_bacteria:*
10: sp_rodent:*
11: sp_virus:*
12: sp_vertebrate:*
13: sp_unclassified:*
14: sp_archea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	820	100.0	184	2	Q15330	Q15330 homo sapien
2	647.5	79.0	184	10	P97682	P97682 rattus norv
3	131	16.0	1574	10	O88281	O88281 rattus norv
4	123.5	15.1	480	2	Q92743	Q92743 homo sapien
5	111	13.5	220	10	Q63404	Q63404 rattus norv
6	113.5	13.8	830	2	Q14162	Q14162 homo sapien
7	113.5	13.8	830	2	O43701	O43701 homo sapien
8	110.5	13.5	381	2	O43775	O43775 homo sapien
9	109	13.3	250	2	O76076	O76076 homo sapien
10	100	12.2	343	12	O42607	O42607 xenopus lae
11	98	12.0	251	3	Q24774	Q24774 enchytraeus
12	97.5	11.9	367	10	O54775	O54775 mus musculu
13	98.5	12.0	1116	12	O73791	O73791 brachydanio
14	97	11.8	1792	12	O57484	O57484 gallus gall
15	96.5	11.8	1645	2	O15230	O15230 homo sapien
16	89.5	10.9	281	10	O88812	O88812 mus musculu
17	95	11.6	1725	10	P70570	P70570 rattus norv
18	92.5	11.3	1816	10	P97927	P97927 mus musculu
19	83	10.1	153	2	O75095	O75095 homo sapien
20	88	10.7	728	12	Q90656	Q90656 gallus gall
21	84.5	10.3	282	2	Q16270	Q16270 homo sapien
22	79.5	9.7	88	8	Q43666	Q43666 triticum ae
23	91.5	11.2	3704	3	P91904	P91904 caenorhabdi
24	85	10.4	547	10	P97881	P97881 rattus norv
25	81.5	9.9	277	2	Q07822	Q07822 homo sapien
26	80	9.8	190	12	O93518	O93518 gloydius ha
27	83	10.1	471	4	O19131	O19131 bos taurus
28	85	10.4	955	2	Q99466	Q99466 homo sapien
29	87.5	10.7	2212	2	Q93072	Q93072 homo sapien
30	82	10.0	530	3	Q24526	Q24526 drosophila
31	78	9.5	169	2	Q14564	Q14564 homo sapien
32	84	10.2	1106	3	Q17494	Q17494 caenorhabdi
33	75.5	9.2	88	8	Q43663	Q43663 triticum ae
34	82	10.0	789	10	P70505	P70505 rattus norv
35	85	10.4	1999	2	Q99940	Q99940 homo sapien
36	85	10.4	2003	2	O00306	O00306 homo sapien
37	80	9.8	452	2	O60295	O60295 homo sapien
38	76.5	9.3	176	3	O16364	O16364 caenorhabdi
39	75	9.1	149	3	O46176	O46176 aedes aegyp
40	80	9.8	727	3	Q94528	Q94528 drosophila
41	78.5	9.6	518	2	Q14810	Q14810 homo sapien
42	78.5	9.6	524	2	Q14808	Q14808 homo sapien
43	79.5	9.7	719	2	O60283	O60283 homo sapien
44	82	10.0	1531	10	O88279	O88279 rattus norv
45	81	9.9	1172	2	O14947	O14947 homo sapien

ALIGNMENTS

```
RESULT   1
Q15330
ID   Q15330      PRELIMINARY;      PRT;   184 AA.
AC   Q15330;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ESM-1 SECRETORY PROTEIN PRECURSOR.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96355375.
RA   LASSALLE P.M., MOLET S., JANIN A., VANDER-HEYDEN J.E., TAVERNIER J.,
RA   FIERS W., DEVOS R.E., TONNEL A.E.B.;
RT   "ESM-1 is a novel human endothelial cell-specific molecule expressed
RT   in lung and regulated by cytokines.";
RL   J. BIOL. CHEM. 271:20458-20464(1996).
DR   EMBL; X89426; E189266; -.
DR   PFAM; PF00219; IGFBP; 1.
KW   SIGNAL.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    184       ESM-1 SECRETORY PROTEIN.
SQ   SEQUENCE   184 AA;  20095 MW;  08D109DF CRC32;

  Query Match             100.0%;  Score 820;  DB 2;  Length 184;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-81;
  Matches  163;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60

Qy      61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120

Qy     121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVR 163
           |||||||||||||||||||||||||||||||||||||||||||
Db     121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVR 163


RESULT   2
P97682
ID   P97682      PRELIMINARY;      PRT;   184 AA.
AC   P97682;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE   PG25.
```

```
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE DAWLEY; TISSUE=PINEAL GLAND;
RA   WANG X., BROWNSTEIN M.J., YOUNG W.S.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U80818; G1750198; -.
SQ   SEQUENCE   184 AA;  20075 MW;  3457564C CRC32;


  Query Match             79.0%;  Score 647.5;  DB 10;  Length 184;
  Best Local Similarity   75.6%;  Pred. No. 1.7e-62;
  Matches  124;  Conservative   21;  Mismatches   18;  Indels    1;  Gaps    1;

Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60
           |||:|||||||:| ||  ||| :||||||:|||::||:|| |||||||||||||:|||||
Db       1 MKSLLLLTTLLIPLHLGMAWSAKYAVDCPEHCDNTECRSSLRCKRTVLDDCGCCQVCAAG 60

Qy      61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120
           ||||||||||||||:||||||:|:  : || ||:|||:||||||||||||||:||||||||
Db      61 PGETCYRTVSGMDGVKCGPGLKCHFYSEEDDFGDEFGVCKDCPYGTFGMDCKETCNCQSG 120

Qy     121 ICDRGTGKCLKFPFFQYSVTKSSNR-FVSLTEHDMASGDGNIVR 163
           |||| ||:|| |||||| :||::|   | ||:| |||||| ||
Db     121 ICDRVTGRCLDFPFFQYAAAKSPSRTSASQTERDAASGDGNAVR 164


RESULT   3
O88281
ID   O88281      PRELIMINARY;      PRT;  1574 AA.
AC   O88281;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MEGF6.
GN   MEGF6.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   NAKAYAMA M., NAKAJIMA D., NAGASE T., NOMURA N., SEKI N., OHARA O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   GENOMICS 51:27-34(1998).
DR   EMBL; AB011532; D1033425; -.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS01186; EGF_2; 23.
DR   PROSITE; PS01187; EGF_CA; 5.
KW   GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ   SEQUENCE   1574 AA;  165445 MW;  C49E6EA6 CRC32;


  Query Match             16.0%;  Score 131;  DB 10;  Length 1574;
  Best Local Similarity   23.9%;  Pred. No. 2.8e-06;
  Matches   37;  Conservative   23;  Mismatches   23;  Indels   72;  Gaps    9;

Qy      15 HLV-AAWSNNYAVDCPQHC--------------DSSECKS---SPRCKRTVLDDC----- 52
           ||  ::|:  ::  |:: |              |  ||:   : ||      :|
Db     642 HLACPPWA--FGPGCSEDCLCEQSHTRSCNPKDGSCSCKAGFQGERCQ----AECESGFF 695

Qy      52 --GCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICK--------- 101
             || : |:                  | ||: |:| :||         |:
Db     696 GPGCRHRCT-----------------CQPGVACDPVSGE---------CRTQCPPGYQG 728

Qy     101 -----DCPYGTFGMDCRETCNCQSGICDRGTGKCL 130
                :|| ||||::|  :|:| :: | | || ||
Db     729 EDCGQECPVGTFGVNCSGSCSCVGAPCHRVTGECL 763


RESULT   4
Q92743
ID   Q92743      PRELIMINARY;      PRT;  480 AA.
AC   Q92743;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NOVEL SERINE PROTEASE.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 97131593.
RA   ZUMBRUNN J., TRUEB B.;
RT   "Primary structure of a putative serine protease specific for
RT   IGF-binding proteins.";
RL   FEBS LETT. 398:187-192(1996).
DR   EMBL; Y07921; E275186; -.
DR   PFAM; PF00050; kazal; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   PFAM; PF00219; IGFBP; 1.
DR   PFAM; PF00595; PDZ; 1.
KW   PROTEASE; SERINE PROTEASE.
SQ   SEQUENCE   480 AA;  51286 MW;  175BB6D9 CRC32;


  Query Match             15.1%;  Score 123.5;  DB 2;  Length 480;
  Best Local Similarity   34.0%;  Pred. No. 6.8e-06;
  Matches   33;  Conservative   18;  Mismatches   25;  Indels   21;  Gaps    5;

Qy       4 VLLLTTLLVPAHLV-AAWSNNYAVDCPQHCDSSECKSSP------RCKRTVLDDCGCCRV 56
           :||| :   :|:|  |::|   |  ||::|::: | : |      | :    | |||| |
Db      12 LLLLLAAPASAQLSRAGRSAPLAAGCPDRCEPARCPPQPEHCEGGRAR----DACGCCEV 67

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFG 93
           |:|  | :|     |::   || ||:|       |||
Db      68 CGAPEGAAC-----GLQEGPCGEGLQCVV-----PFG 94


RESULT   5
Q63404
ID   Q63404      PRELIMINARY;      PRT;  220 AA.
AC   Q63404;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   (CLONE REM4) ORF (FRAGMENT).
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HOLZMAN; TISSUE=BRAIN;
RX   MEDLINE; 96235155.
RA   ASAKURA K., POGULIS R.J., PEASE L.R., RODRIGUEZ M.;
RT   "A monoclonal autoantibody which promotes central nervous system
RT   remyelination is highly polyreactive to multiple known and novel
RT   antigens.";
RL   J. NEUROIMMUNOL. 65:11-19(1996).
DR   EMBL; L41686; G780367; -.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PFAM; PF00008; EGF; 3.
KW   GLYCOPROTEIN.
FT   NON_TER       1      1
FT   NON_TER     220    220
SQ   SEQUENCE   220 AA;  23231 MW;  4389BC39 CRC32;


  Query Match             13.5%;  Score 111;  DB 10;  Length 220;
  Best Local Similarity   28.0%;  Pred. No. 7.9e-05;
```

```
  Matches   35;  Conservative   24;  Mismatches   40;  Indels   26;  Gaps    9;

Qy      21 SNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGC--------C-RVCAAGR-GETCYRTVS 70
           : : : :| |:|    | ::  | : |  :|:|        | : |: || |  | :
Db      29 PGKHGPQCEQRC---PCQNGGVCHH-VTGECSCPSGWMGTVCGQPCPEGRFGKNCSQECQ 84

Qy      71 GMDGMKCGPGL---RCQPSNGEDPFGEEFGICKD-CPYGTFGMDCRETCNC-QSGICDRG 125
             :|  | ::    :| |:   :        | | || ||:|: | ||| |  :| | :
Db      85 CHNGGACDAATGQCHCSPGYTGER-------CQDECPVGTYGVRCAETCRCVNGGKCYHV 137

Qy     126 TGKCL 130
           :| ||
Db     138 SGTCL 142


RESULT   6
Q14162
ID   Q14162      PRELIMINARY;      PRT;   830 AA.
AC   Q14162;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   KIAA0149 PROTEIN.
GN   KIAA0149.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96127530.
RA   NAGASE T., SEKI N., TANAKA A., ISHIKAWA K., NOMURA N.;
RT   "Prediction of the coding sequences of unidentified human genes. IV.
RT   The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA RES. 2:167-174(1995).
DR   EMBL; D63483; G1469880; -.
DR   PROSITE; PS01186; EGF_2; 6.
KW   GLYCOPROTEIN.
SQ   SEQUENCE   830 AA;  87460 MW;  CC490A76 CRC32;


  Query Match             13.8%;  Score 113.5;  DB 2;  Length 830;
  Best Local Similarity   24.4%;  Pred. No. 0.00013;
  Matches   44;  Conservative   17;  Mismatches   32;  Indels   87;  Gaps   10;

Qy      18 AAWSNNYAVDCPQHCD---------SSECKSSP-------------------------RC 43
           ::|    : :| |:|:         |:||  :|                          ||
Db     212 PGW---WGPECQQQCECVRGRCSAASGECTCPPGFRGARCELPCPAGSHGVQCAHSCGRC 268

Qy      44 KRTVLDDC----GCCRVCAAGR--------------GETCYRTVSGMDGMKCGPGLRCQP 85
           |:   : |    | |  | :|:              ||:| :  :      |  |  |:|
Db     269 KHN--EPCSPDTGSCESCEPGWNGTQCQQPCLPGTFGESCEQQCP-----HCRHGEACEP 321

Qy      86 SNG--------------EDPFGEEFGICKDCPYGTFGMDCRETC-NCQSGICDRGTGKCL 130
             |              |||           || |||| ||  ||  |  | ||  || |:
Db     322 DTGHCQRCDPGWLGPRCEDP----------CPTGTFGEDCGSTCPTCVQGSCDTVTGDCV 371


RESULT   7
O43701
ID   O43701      PRELIMINARY;      PRT;   830 AA.
AC   O43701;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ACETYL LDL RECEPTOR PRECURSOR.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98058897.
RA   ADACHI H., TSUJIMOTO M., ARAI H., INOUE K.;
RT   "Expression cloning of a novel scavenger receptor from human
RT   endothelial cells.";
RL   J. BIOL. CHEM. 272:31217-31220(1997).
DR   EMBL; D86864; D1024977; -.
DR   PROSITE; PS01186; EGF_2; 6.
KW   SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        16    830       ACETYL LDL RECEPTOR.
SQ   SEQUENCE   830 AA;  87430 MW;  936CC3DD CRC32;


  Query Match             13.8%;  Score 113.5;  DB 2;  Length 830;
  Best Local Similarity   24.4%;  Pred. No. 0.00013;
  Matches   44;  Conservative   17;  Mismatches   32;  Indels   87;  Gaps   10;

Qy      18 AAWSNNYAVDCPQHCD---------SSECKSSP-------------------------RC 43
           ::|    : :| |:|:         |:||  :|                          ||
Db     212 PGW---WGPECQQQCECVRGRCSAASGECTCPPGFRGARCELPCPAGSHGVQCAHSCGRC 268

Qy      44 KRTVLDDC----GCCRVCAAGR--------------GETCYRTVSGMDGMKCGPGLRCQP 85
           |:   : |    | |  | :|:              ||:| :  :      |  |  |:|
Db     269 KHN--EPCSPDTGSCESCEPGWNGTQCQQPCLPGTFGESCEQQCP-----HCRHGEACEP 321

Qy      86 SNG--------------EDPFGEEFGICKDCPYGTFGMDCRETC-NCQSGICDRGTGKCL 130
             |              |||           || |||| ||  ||  |  | ||  || |:
Db     322 DTGHCQRCDPGWLGPRCEDP----------CPTGTFGEDCGSTCPTCVQGSCDTVTGDCV 371


RESULT   8
O43775
ID   O43775      PRELIMINARY;      PRT;   381 AA.
AC   O43775;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE   CYR61 PROTEIN.
GN   CYR61.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MARTINERIE C., VIEGAS-PEQUIGNOT E., NGUYEN V.C., PERBAL B.;
RL   J. CLIN. PATHOL. MOL. PATHOL. 50:130-136(1997).
DR   EMBL; Y11307; E304665; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   381 AA;  42025 MW;  1B18FF1A CRC32;


  Query Match             13.5%;  Score 110.5;  DB 2;  Length 381;
  Best Local Similarity   26.1%;  Pred. No. 0.00014;
  Matches   36;  Conservative   21;  Mismatches   44;  Indels   37;  Gaps    8;

Qy       2 KSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRT---VLDDCGCCRVCA 58
           ::: |: |||  ::|  :         ||  |  :    :|:|      | | ||||:|||
Db       7 RALALVVTLLHLTRLALS-------TCPAACHCP--LEAPKCAPGVGLVRDGCGCCKVCA 57

Qy      59 AGRGETCYRTVSGMDGMKCGP--GLRCQPSNGEDPFGEEFGICKD------CPYG----- 106
             | | :| :      |    || |  : :  ::    |||:       | |
Db      58 KQLNEDCSKTQP------CDHTKGLECNFGASSTALK---GICRAQSEGRPCEYNSRIYQ 108

Qy     106 ---TFGMDCRETCNCQSG 120
              :|  :|:  | |  |
Db     109 NGESFQPNCKHQCTCIDG 126


RESULT   9
O76076
ID   O76076      PRELIMINARY;      PRT;   250 AA.
AC   O76076;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
```

```
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CONNECTIVE TISSUE GROWTH FACTOR-LIKE PROTEIN PRECURSOR.
GN   CT58.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   KUMAR S., HAND A.T., CONNOR J.C., DODDS R.A., RYAN P.J., TRILL J.G.,
RA   FISHER S.M., SLEMMON J.R., LIPSHUTZ D.B., BARTHOLOMEW V., JAMES I.E.,
RA   RIEMAN D.J., GOWEN M., LEE J.C.;
RT   "Identification and cloning of CTGF-L from human osteoblasts, a novel
RT   cysteine rich protein containing an IGF binding domain.";
RL   J. BONE MINER. RES. 13:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   ROWLES J., GENDLER S.;
RT   "CT58, a new member of the connective tissue growth factor family,
RT   interacts with the breast cancer associated mucin MUC1.";
RL   SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF083500; G3462836; -.
DR   EMBL; AF074604; G3328192; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
SQ   SEQUENCE   250 AA;  26825 MW;  61CE7682 CRC32;
```

```
  Query Match             13.3%;  Score 109;  DB 2;  Length 250;
  Best Local Similarity   30.4%;  Pred. No. 0.00014;
  Matches   42;  Conservative   16;  Mismatches   44;  Indels   36;  Gaps    8;

Qy       2 KSVLLLTTLL-----VPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGC 53
           |: ||  :||     | ::|           ||  |  :   ::|||      ||| |||
Db       6 KTHLLAFSLLCLLSKVRTQL----------CPTPCTCP--WPPPRCPLGVPLVLDGCGC 52

Qy      54 CRVCAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDP------FGEEFGIC--KDCPY- 105
           |||||   || |      :     : || |||: |         ::|: : |  :   |
Db      53 CRVCARRLGEPC----DQLHVCDASQGLVCQPGAGPGGRGALCLLAEDDSSCEVNGRLYR 108

Qy     105 --GTFGMDCRETCNCQSG 120
             ||   |   | |: |
Db     109 EGETFQPHCSIRCRCEDG 126
```

```
RESULT  10
O42607
ID   O42607      PRELIMINARY;      PRT;   343 AA.
AC   O42607;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CONNECTIVE TISSUE GROWTH FACTOR XCTGF.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YING Z., KING M.L.;
RL   SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U43524; G2342756; -.
DR   EMBL; U43523; G2342754; -.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   PFAM; PF00007; Cys_knot; 1.
DR   PFAM; PF00090; tsp_1; 1.
DR   PFAM; PF00093; vwc; 1.
DR   PFAM; PF00219; IGFBP; 1.
SQ   SEQUENCE   343 AA;  37966 MW;  4F7E7114 CRC32;
```

```
  Query Match             12.2%;  Score 100;  DB 12;  Length 343;
  Best Local Similarity   27.3%;  Pred. No. 0.0017;
  Matches   39;  Conservative   18;  Mismatches   38;  Indels   48;  Gaps   11;

Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVC 57
           : :|||:: :          | :: | :|    |: :  :  | |    | | | ||||:||
Db       6 VTAVLLFALF--------CWVSD-AQECNGECQCP--NKVPVCDPGVRMVQDGCGCCKVC 54

Qy      58 AAGRGETCY-RTVSGMDGMKCGP--GLRCQPSNGEDPFGE----EFGICKD-----CPYG 105
           :   || |  | |        | |  || |:       ||       |:|       | :|
Db      55 SKQLGELCTERDV-------CDPHKGLFCD-------FGSRVNRKIGVCTAREGAPCVFG 100

Qy     106 --------TFGMDCRETCNCQSG 120
                   :|   |:  | | |
Db     101 GTVYRSGESFQSSCKYQCTCIDG 123
```

```
RESULT  11
Q24774
ID   Q24774      PRELIMINARY;      PRT;   251 AA.
AC   Q24774;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CYSTEINE-RICH-PROTEIN.
GN   CRP.
OS   ENCHYTRAEUS BUCHHOLZI.
OC   EUKARYOTA; METAZOA; ANNELIDA; CLITELLATA; OLIGOCHAETA; HAPLOTAXIDA;
OC   TUBIFICINA; ENCHYTRAEIDAE; ENCHYTRAEUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95014230.
RA   WILLUHN J., SCHMITT-WREDE H.P., GREVEN H., WUNDERLICH F.;
RT   "cDNA cloning of a cadmium-inducible mRNA encoding a novel
RT   cysteine-rich, non-metallothionein 25-kDa protein in an enchytraeid
RT   earthworm.";
RL   J. BIOL. CHEM. 269:24688-24691(1994).
DR   EMBL; X79344; G488803; -.
SQ   SEQUENCE   251 AA;  25000 MW;  85747E80 CRC32;
```

```
  Query Match             12.0%;  Score 98;  DB 3;  Length 251;
  Best Local Similarity   25.4%;  Pred. No. 0.0022;
  Matches   32;  Conservative   18;  Mismatches   34;  Indels   42;  Gaps    7;

Qy      32 CDSSECK-------SSPRCKRTVLDDCGCCRVCAAGRGE-------TCYRTVSGMDGMKC 77
           |: :|||        :|:|    :  |:|   |   :||        |  |  |::   |
Db      81 CEKGECKKGCKEGCCAPKC---GVAGCSCGSGCKCEKGECKPGCTKRCCGTKCGVEDCPC 137

Qy      78 GPGLRCQPSNGEDPFGEEFGICK-DCP-------YGTFGMDCRETCNCQSGICDRG---- 126
           ||: :|: ::          || :|:       |  |  |   | |: | | :|
Db     138 GPSCKCEKGD----------CKVNCSKGCCTPKCGVVGCPCGSQCTCEKGSCKKGCSKG 186

Qy     126 --TGKC 129
               | ||
Db     187 CCTPKC 192
```

```
RESULT  12
O54775
ID   O54775      PRELIMINARY;      PRT;   367 AA.
AC   O54775;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ELM1.
GN   ELM1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HEN;
RX   MEDLINE; 98119879.
RA   HASHIMOTO Y., SHINDO-OKADA N., TANI M., NAGAMACHI Y., TAKEUCHI K.,
RA   SHIROISHI T., TOMA H., YOKOTA J.;
```

```
RT   "Expression of the Elm1 gene, a novel gene of the CCN (connective
RT   tissue growth factor, Cyr61/Cef10, and neuroblastoma overexpressed
RT   gene) family, suppresses In vivo tumor growth and metastasis of K-1735
RT   murine melanoma cells.";
RL   J. EXP. MED. 187:289-296(1998).
DR   EMBL; AB004873; D1025874; -.
DR   PROSITE; PS01185; CTCK_1; 1.
SQ   SEQUENCE   367 AA;  40702 MW;  1AB35AB9 CRC32;


  Query Match             11.9%;  Score 97.5;  DB 10;  Length 367;
  Best Local Similarity   28.2%;  Pred. No. 0.0034;
  Matches   35;  Conservative   15;  Mismatches   35;  Indels   39;  Gaps   10;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           :|||        |:|||      : | | ||::||   |: |          ::  | |  ||
Db      54 ECPQ--------SPPRCPLGVSLITDGCECCKICAQQLGDNC------TEAAICDPHRGL 99

Qy      82 RCQPSNGEDPFGEEFGICKD-----C-----PY---GTFGMDCRETCNCQSGI--CDRGT 126
           |: |     ::   |:| :     |       |    :|  :||  | |  |   |   |
Db     100 YCDYSGDRPRYA--IGVCAQVVGVGCVLDGVRYTNGESFQPNCRYNCTCIDGTVGC---T 154

Qy     127 GKCL 130
           ||
Db     155 PLCL 158


RESULT  13
O73791
ID   O73791      PRELIMINARY;      PRT;  1116 AA.
AC   O73791;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   ENDOTHELIUM-SPECIFIC RECEPTOR TYROSINE KINASE TIE-2.
GN   TIE-2.
OS   BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC   CYPRINIDAE; RASBORINAE; DANIO.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LYONS M.S., BELL B., STAINIER D., PETERS K.G.;
RL   DEV. DYN. 0:0-0(1998).
DR   EMBL; AF053632; G3005607; -.
SQ   SEQUENCE   1116 AA;  122361 MW;  7E60F6C5 CRC32;


  Query Match             12.0%;  Score 98.5;  DB 12;  Length 1116;
  Best Local Similarity   23.7%;  Pred. No. 0.0066;
  Matches   36;  Conservative   18;  Mismatches   33;  Indels   65;  Gaps    8;

Qy       7 LTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG-RGETC 65
           :| |:| :   : |: | :  ||:  ::: |  :          :|    ||::| || ||
Db     207 ITRLIVRSCRAGFWGPNCTESCPRCANGGVCDDTT-------GEC----VCPPGFRGHTC 255

Qy      66 YRTVSGMDGMKCGPG--------------------LRCQPSNGEDPFGEEFGIC----- 100
                 : || |                       ||      ||:|     |
Db     256 --------DIVCGEGRFGAGCKERCVDGVCRALVFCLR-------DPYG-----CSCASG 295

Qy     100 -------KDCPYGTFGMDCRETCNCQSGICDR 124
                  || | :|  | : | |  | |||
Db     296 WRGLSCNDACPDGYYGAGCTQKCVCAKGRCDR 327


RESULT  14
O57484
ID   O57484      PRELIMINARY;      PRT;  1792 AA.
AC   O57484;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   LAMININ BETA 2-LIKE CHAIN.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93015947.
RA   O'REAR J.J.;
RT   "A novel laminin B1 chain variant in avian eye.";
RL   J. BIOL. CHEM. 267:20555-20557(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   LIU J., SWASDISON S., XIE W., BREWTON R.G., MAYNE R.;
RL   MATRIX BIOL. 16:0-0(1998).
DR   EMBL; AF038555; G2708707; -.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
KW   GLYCOPROTEIN.
SQ   SEQUENCE   1792 AA;  195722 MW;  AFFC9020 CRC32;


  Query Match             11.8%;  Score 97;  DB 12;  Length 1792;
  Best Local Similarity   23.0%;  Pred. No. 0.014;
  Matches   43;  Conservative   22;  Mismatches   36;  Indels   86;  Gaps   12;

Qy      29 PQHCDSSECKS----SPRCKRTVLDDCGC--------------------CRVCA-AGRG 62
           |::||: :| :    |: |  ||  :| |                    || |   |::
Db     820 PRRCDPCQCSGEGSLSAVCD-TVTGQCPCRQGAHGSRCDRCQPGHWGFPTCRPCQCNGHA 878

Qy      63 E-------TCYRTVSGMDGMKC-------------GPGLRCQPSN-GEDPFGE-EFGI-- 99
           |          :| |     || :|             |:| :|:|    : | |   |:
Db     879 EECDPQTGSCLRCRDHTDGERCQRCAAGHFGNPALGSGQHCRPCPCPDGPGGPRHFAASC 938

Qy      99 --------------------CKDCPYGTF------GMDCRETCNCQSGI-------CDR 124
                               | :|: | :      |  ||  | |:: |       |||
Db     939 YQDGRSRQVVCHCSPGYTGPRCDECAPGYYGDPLQPGGHCR-PCQCHNNIDTTDPEACDR 997

Qy     125 GTGKCLK 131
           || ||:
Db     998 RTGQCLR 1004


RESULT  15
O15230
ID   O15230      PRELIMINARY;      PRT;  1645 AA.
AC   O15230;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   KIAA0533 PROTEIN (LAMININ ALPHA 5 CHAIN) (FRAGMENT).
GN   KIAA0533 OR LAMA5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98290545.
RA   NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,
RA   OHARA O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA RES. 5:31-39(1998).
RN   [2]
RP   SEQUENCE OF 693-1645 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 97415425.
RA   DURKIN M.E., LOECHEL F., MATTEI M., GILPIN B.J., ALBRECHTSEN R.,
RA   WEWER U.M.;
RT   "Tissue-specific expression of the human laminin alpha5-chain, and
RT   mapping of the gene to human chromosome 20q13.2-13.3 and to distal
RT   mouse chromosome 2 near the locus for the ragged (Ra) mutation.";
RL   FEBS LETT. 411:296-300(1997).
```

```
DR   EMBL; AB011105; D1026389; -.
DR   EMBL; Z95636; E317479; -.
DR   PFAM; PF00054; laminin_G; 2.
FT   NON_TER       1       1
SQ   SEQUENCE   1645 AA;  178215 MW;  A44DE58F CRC32;
```

Query Match 11.8%; Score 96.5; DB 2; Length 1645;
Best Local Similarity 30.1%; Pred. No. 0.015;
Matches 28; Conservative 12; Mismatches 20; Indels 33; Gaps 6;

```
Qy      48 LDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP---GLRCQPSNGE-----DPFGEEFGIC 99
           :: || || ||                  |||   |  |:| :|:        | :   |
Db      11 FNGCGGCRPCA-----------------CGPAAEGSECHPQSGQCHCRPGTMGPQ---C 49

Qy     100 KDCPYGTFGM---DCRETCNCQSGICDRGTGKC 129
           ::|: |  |:    ||  | | :| ||  ||:|
Db      50 RECAPGYWGLPEQGCR-RCQCPGGRCDPHTGRC 81
```

Search completed: May 3, 1999, 17:54:22
Job time: 642 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 14:59:01 ; Search time 18.06 Seconds
(without alignments)
242.246 Million cell updates/sec

Title: US-09-037-460-2_COPY_1_163
Perfect score: 820
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....NRFVSLTEHDMASGDGNTVR 163

Scoring table: PAM150

Searched: 74019 seqs, 26840295 residues

Database : SwissProt_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	141.5	17.3	351	1	NOV_CHICK	P28686 gallus gall
2	139	17.0	353	1	NOV_COTJA	P42642 coturnix co
3	124.5	15.2	354	1	NOV_MOUSE	Q64299 mus musculu
4	123.5	15.1	357	1	NOV_HUMAN	P48745 homo sapien
5	115.5	14.1	258	1	IBP4_BOVIN	Q05716 bos taur
6	114.5	14.0	271	1	IBP5_MOUSE	Q07079 mus musculu
7	112	13.7	348	1	CTGF_MOUSE	P29268 mus musculu
8	110.5	13.5	271	1	IBP5_RAT	P24594 rattus norv
9	111	13.5	375	1	CE10_CHICK	P19336 gallus gall
10	110.5	13.5	381	1	CYR6_HUMAN	O00622 homo sapien
11	109.5	13.4	272	1	IBP5_HUMAN	P24593 homo sapien
12	109	13.3	349	1	CTGF_HUMAN	P29279 homo sapien
13	107	13.0	237	1	IBP4_SHEEP	Q28893 ovis aries
14	106.5	13.0	258	1	IBP4_HUMAN	P22692 homo sapien
15	104.5	12.7	266	1	IBP3_PIG	P16611 sus scrofa
16	104.5	12.7	271	1	IBP5_PIG	Q28985 sus scrofa
17	103.5	12.6	254	1	IBP4_MOUSE	P47879 mus musculu
18	103.5	12.6	254	1	IBP4_RAT	P21744 rattus norv
19	103.5	12.6	379	1	CYR6_MOUSE	P18406 mus musculu
20	103	12.6	349	1	CTGF_PIG	O19113 sus scrofa
21	101.5	12.4	291	1	IBP3_BOVIN	P20959 bos taurus
22	101.5	12.4	291	1	IBP3_HUMAN	P17936 homo sapien
23	100.5	12.3	291	1	IBP3_MOUSE	P47878 mus musculu
24	99	12.1	343	1	NOV_XENLA	P51609 xenopus lae
25	98.5	12.0	292	1	IBP3_RAT	P15473 rattus norv
26	103	12.6	1801	1	LMB2_RAT	P15800 rattus norv
27	101.5	12.4	1138	1	TIE1_HUMAN	P35590 homo sapien
28	97	11.8	317	1	IBP2_SHEEP	Q29400 ovis aries
29	94.5	11.5	259	1	IBP1_HUMAN	P08833 homo sapien
30	100	12.2	1799	1	LMB2_MOUSE	Q61292 mus musculu
31	94	11.5	311	1	IBP2_CHICK	P49705 gallus gall
32	97.5	11.9	1136	1	TIE1_BOVIN	Q06805 bos taurus
33	94.5	11.5	1134	1	TIE1_MOUSE	Q06806 mus musculu
34	87.5	10.7	111	1	IBP5_BOVIN	Q05717 bos taurus
35	94.5	11.5	1786	1	LMB1_MOUSE	P02469 mus musculu
36	95.5	11.6	3635	1	LMA5_MOUSE	Q61001 mus musculu
37	92	11.2	1786	1	LMB1_HUMAN	P07942 homo sapien
38	91.5	11.2	1816	1	LMA4_HUMAN	Q16363 homo sapien
39	86.5	10.5	328	1	IBP2_HUMAN	P18065 homo sapien
40	94	11.5	5147	1	FAT_DROME	P33450 drosophila
41	86	10.5	349	1	CTGF_BOVIN	O18739 bos taurus
42	90	11.0	1639	1	LMG1_DROME	P15215 drosophila
43	84.5	10.3	304	1	IBP2_RAT	P12843 rattus norv
44	84.5	10.3	305	1	IBP2_MOUSE	P47877 mus musculu
45	91.5	11.2	3672	1	LML2_CAEEL	Q21313 caenorhabdi

ALIGNMENTS

```
RESULT  1
NOV_CHICK
ID   NOV_CHICK      STANDARD;      PRT;   351 AA.
AC   P28686;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN PRECURSOR.
GN   NOV.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BROWN LEGHORN;
RX   MEDLINE; 92107157.
RA   MALOISEL V., MARTINERIE C., DAMBRINE G., PLASSIART G., BRISAC M.,
RA   CROCHET J., PERBAL B.;
RL   MOL. CELL. BIOL. 12:10-21(1992).
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION. ITS OVEREXPRESSION IS ASSOCIATED WITH
CC       TUMORIGENESIS AND EXPRESSION OF A N-TERMINAL-TRUNCATED VERSION
CC       OF NOV GENE IN CHICKEN EMBRYONIC FIBROBLASTS (CEF) IS SUFFICIENT
CC       TO INDUCE THE TRANSFORMATION OF CEF IN VITRO.
CC   -!- TISSUE SPECIFICITY: BRAIN AND HEART, AND AT A LOWER LEVEL IN
CC       MUSCLE AND INTESTINE, IN THE EMBRYO. LUNG AND LESS SO IN BRAIN AND
CC       SPLEEN, IN ADULT CHICKEN.
CC   -!- DEVELOPMENTAL STAGE: MAV1-INDUCED NEPHROBLASTOMAS EXPRESS A HIGH
CC       LEVEL OF NOV GENE WHOSE TRANSCRIPTION IS NORMALLY ARRESTED IN
CC       ADULT KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; X59284; G63703; -.
DR   PIR; S20078; S20078.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    351       NOV PROTEIN.
FT   DOMAIN      104    170       VWFC.
FT   DOMAIN      258    332       CTCK.
FT   DISULFID    258    295       BY SIMILARITY.
FT   DISULFID    275    309       BY SIMILARITY.
FT   DISULFID    286    325       BY SIMILARITY.
FT   DISULFID    289    327       BY SIMILARITY.
FT   DISULFID    294    331       BY SIMILARITY.
FT   CARBOHYD    274    274       POTENTIAL.
SQ   SEQUENCE   351 AA;  38268 MW;  C7044065 CRC32;
```

Query Match 17.3%; Score 141.5; DB 1; Length 351;
Best Local Similarity 31.9%; Pred. No. 2.3e-08;
Matches 45; Conservative 17; Mismatches 38; Indels 41; Gaps 7;

```
Qy       3 SVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAA 59
           :|||| ||: :  |:::      ||: | :   :|||    :||| |||| |||
Db      10 PVLLLLLLLLRPCEVSGR----EAACPRPCGGRCPAEPPRCAPGVPAVLDGCGCCLVCAR 65

Qy      60 GRGETCYRTVSGMDGMKCGPGLRCQPSNG-------EDPFGEEFGICK-----DCPYG-- 106
            |||:           |:| | |: | |       || |   |||      :| :
Db      66 QRGES-----------CSPLLPCDESGGLYCDRGPED--GGGAGICMVLEGDNCVFDGM 111
```

```
Qy      106 ------TFGMDCRETCNCQSG 120
            |||||||||||||||||||||
Db      112 IYRNGETFQPSCKYQCTCRDG 132
```

```
RESULT   2
NOV_COTJA
ID   NOV_COTJA       STANDARD;       PRT;   353 AA.
AC   P42642;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN PRECURSOR.
GN   NOV.
OS   COTURNIX COTURNIX JAPONICA (JAPANESE QUAIL).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WEISKIRCHEN R., BISTER K.;
RL   SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; U13063; G532697; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    353       NOV PROTEIN.
FT   DOMAIN      106    172       VWFC.
FT   DOMAIN      260    334       CTCK.
FT   DISULFID    260    297       BY SIMILARITY.
FT   DISULFID    277    311       BY SIMILARITY.
FT   DISULFID    288    327       BY SIMILARITY.
FT   DISULFID    291    329       BY SIMILARITY.
FT   DISULFID    296    333       BY SIMILARITY.
FT   CARBOHYD    276    276       POTENTIAL.
SQ   SEQUENCE   353 AA;  38667 MW;  C4F5928D CRC32;
```

```
  Query Match             17.0%;  Score 139;  DB 1;  Length 353;
  Best Local Similarity   32.1%;  Pred. No. 4.4e-08;
  Matches   45;  Conservative   14;  Mismatches   39;  Indels   42;  Gaps    7;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAG 60
           :|||  || |: :    |     ||: | :    :|||     :||| |||| |||
Db      14 LLLLLLLLRPSEV-----NGREAPCPRPCGGRCPAEPPRCAPGVPAVLDGCGCCLVCARQ 68

Qy      61 RGETCYRTVSGMDGMKCGPGLRCQPSNG-------EDPFGEEFGICK-----DCPYG--- 106
           |||:           |:| | |: | |       || |   |||      :| :
Db      69 RGES-----------CSPLLPCDESGGLYCDRGPED--GGGTGICMVLEGDNCVFDGMI 114

Qy     106 -----TFGMDCRETCNCQSG 120
                ||   |:  | |: |
Db     115 YRNGETFQPSCKYQCTCRDG 134
```

```
RESULT   3
NOV_MOUSE
ID   NOV_MOUSE       STANDARD;       PRT;   354 AA.
AC   Q64299;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN HOMOLOG PRECURSOR (NOVH).
GN   NOV.
OS   MUS MUSCULUS (MOUSE).
```

```
RA   SNAITH M.R., NATARAJAN D.; TAYLOR L.B., CHOI C.P., MARTINERIE C.,
RA   PERBAL B., SCHOFIELD P.N., BOULTER C.A.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE; 96204003.
RA   MARTINERIE C., CHEVALIER G., RAUSCHER F.J. III, PERBAL B.;
RL   ONCOGENE 12:1479-1492(1996).
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; X97863; E1181580; -.
DR   EMBL; Y09257; E281106; -.
DR   EMBL; X96585; E228599; -.
DR   MGD; MGI:109185; NOV.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    354       NOV PROTEIN HOMOLOG.
FT   DOMAIN      102    168       VWFC.
FT   DOMAIN      261    335       CTCK.
FT   DISULFID    261    298       BY SIMILARITY.
FT   DISULFID    278    312       BY SIMILARITY.
FT   DISULFID    289    328       BY SIMILARITY.
FT   DISULFID    292    330       BY SIMILARITY.
FT   DISULFID    297    334       BY SIMILARITY.
FT   CARBOHYD     91     91       POTENTIAL.
FT   CARBOHYD    277    277       POTENTIAL.
SQ   SEQUENCE   354 AA;  38928 MW;  65A4FF0E CRC32;
```

```
  Query Match             15.2%;  Score 124.5;  DB 1;  Length 354;
  Best Local Similarity   32.7%;  Pred. No. 1.6e-06;
  Matches   37;  Conservative   18;  Mismatches   27;  Indels   31;  Gaps    9;

Qy      28 CPQHCDSSECKS-SPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMK-C--GPG 80
           || :| :: | | || |    |:||| |:|| |||  |||:|          |: |  ::|
Db      29 CPSRC-PPKCPSISPTCAPGVRSVLDGCSCCPVCARQRGESC-------SEMRPCDQSSG 80

Qy      81 LRCQPSNGEDPFGEEFGICK-----DCPYG--------TFGMDCRETCNCQSG 120
           | |: |    :   :: |||      :| :          | :|:  | |: |
Db      81 LYCDRSADPN---NQTGICMVPEGDNCVFDGVIYRNGEKFEPNCQYFCTCRDG 130
```

```
RESULT   4
NOV_HUMAN
ID   NOV_HUMAN       STANDARD;       PRT;   357 AA.
AC   P48745;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN HOMOLOG PRECURSOR (NOVH).
GN   NOV.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 94336229.
RA   MARTINERIE C., HUFF V., JOUBERT I., BADZIOCH M., SAUNDERS G.,
```

```
RA   STRONG L., PERBAL B.;
RL   ONCOGENE 9:2729-2732(1994).
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: INCREASED EXPRESSION IN WILMS TUMOR OF THE
CC       STROMAL TYPE.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; X78351; G825696; -.
DR   EMBL; X78352; G825696; JOINED.
DR   EMBL; X78353; G825696; JOINED.
DR   EMBL; X78354; G825696; JOINED.
DR   EMBL; X96584; E228691; -.
DR   MIM; 164958; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    357       NOV PROTEIN HOMOLOG.
FT   DOMAIN      108    174       VWFC.
FT   DOMAIN      264    338       CTCK.
FT   DISULFID    264    301       BY SIMILARITY.
FT   DISULFID    281    315       BY SIMILARITY.
FT   DISULFID    292    331       BY SIMILARITY.
FT   DISULFID    295    333       BY SIMILARITY.
FT   DISULFID    300    337       BY SIMILARITY.
FT   CARBOHYD     97     97       POTENTIAL.
FT   CARBOHYD    280    280       POTENTIAL.
SQ   SEQUENCE   357 AA;  39162 MW;  DA8B009D CRC32;


  Query Match             15.1%;  Score 123.5;  DB 1;  Length 357;
  Best Local Similarity   31.8%;  Pred. No. 2.1e-06;
  Matches   42;  Conservative   25;  Mismatches   34;  Indels   31;  Gaps    9;

Qy       7 LTTLLVPAHLVA--AWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGR 61
           || ||:  ||::  | :   : :|| :| :    ::| |    |:||| |:|| |||  |
Db      18 LTFLLL--HLLGQVAATQRCPPQCPGRCPA----TPPTCAPGVRAVLDGCSCCLVCARQR 71

Qy      62 GETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICK-----DCPYG--------TFG 108
           ||:|    | ::    ::|| |: |   || ::: |||      :| :          |
Db      72 GESC----SDLEPCDESSGLYCDRS--ADP-SNQTGICTAVEGDNCVFDGVIYRSGEKFQ 124

Qy     109 MDCRETCNCQSG 120
             |:  | |: |
Db     125 PSCKFQCTCRDG 136


RESULT   5
IBP4_BOVIN
ID   IBP4_BOVIN     STANDARD;      PRT;   258 AA.
AC   Q05716;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 PRECURSOR (IGFBP-4)
DE   (IBP-4) (IGF-BINDING PROTEIN 4).
GN   IGFBP4.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93125553.
RA   MOSER D.R., LOWE W.L. JR., DAKE B.L., BOOTH B.A., BOES M.,
RA   CLEMMONS D.R., BAR R.S.;
RL   MOL. ENDOCRINOL. 6:1805-1814(1992).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- BINDS IGF-II MORE THAN IGF-I.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; S52770; G263304; -.
DR   PIR; A45403; A45403.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     21       BY SIMILARITY.
FT   CHAIN        22    258       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN 4.
FT   CARBOHYD    125    125       POTENTIAL.
FT   DOMAIN      200    249       THYROGLOBULIN TYPE I.
SQ   SEQUENCE   258 AA;  27890 MW;  97880748 CRC32;


  Query Match             14.1%;  Score 115.5;  DB 1;  Length 258;
  Best Local Similarity   32.3%;  Pred. No. 1.2e-05;
  Matches   30;  Conservative   24;  Mismatches   30;  Indels    9;  Gaps    4;

Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVC 57
           | |: |:::||::|   :: ::: |: ||   :   : |:::  |   | :  ||||  |
Db       1 MLSLCLMAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATC 59

Qy      58 AAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           | |:|  |     |:   :||:|||| |: | :
Db      60 ALGKGMPC-----GVYTPRCGSGLRCYPPRGVE 87


RESULT   6
IBP5_MOUSE
ID   IBP5_MOUSE     STANDARD;      PRT;   271 AA.
AC   Q07079;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PRECURSOR (IGFBP-5)
DE   (IBP-5) (IGF-BINDING PROTEIN 5).
GN   IGFBP5 OR IGFBP-5.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=MYOBLASTS;
RX   MEDLINE; 94042976.
RA   JAMES P.L., JONES S.B., BUSBY W.H. JR., CLEMMONS D.R., ROTWEIN P.;
RL   J. BIOL. CHEM. 268:22305-22312(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPLEEN;
RX   MEDLINE; 94307727.
RA   KOU K., JENKINS N.A., GILBERT D.J., COPELAND N.G., ROTWEIN P.;
RL   GENOMICS 20:412-418(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 95121750.
RA   SCHULLER A.G.P., GROFFEN C., VAN NECK J.W., ZWARTHOFF E.C.,
RA   DROP S.L.S.;
RL   MOL. CELL. ENDOCRINOL. 104:57-66(1994).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, UTERUS AND
CC       GASTROCNEMIUS MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; L12447; G293384; -.
```

```
DR   EMBL; U02025; G437125; -.
DR   EMBL; U02023; G437125; JOINED.
DR   EMBL; U02027; G437125; JOINED.
DR   EMBL; U02024; G437125; JOINED.
DR   EMBL; X81583; G550385; -.
DR   MGD; MGI:96440; IGFBP5.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    271       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN 5.
FT   DOMAIN      214    262       THYROGLOBULIN TYPE I.
FT   CONFLICT    112    112       MISSING (IN REF. 2).
SQ   SEQUENCE   271 AA;  30372 MW;  12DC64CA CRC32;


  Query Match             14.0%;  Score 114.5;  DB 1;  Length 271;
  Best Local Similarity   28.7%;  Pred. No. 1.6e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAAYAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT   7
CTGF_MOUSE
ID   CTGF_MOUSE     STANDARD;      PRT;   348 AA.
AC   P29268;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CONNECTIVE TISSUE GROWTH FACTOR PRECURSOR (CTGF) (FISP-12 PROTEIN).
GN   CTGF OR FISP12 OR FISP-12.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91363290.
RA   RYSECK R.-P., MACDONALD-BRAVO H., MATTEI M.-G., BRAVO R.;
RL   CELL GROWTH DIFFER. 2:225-233(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91229699.
RA   BRUNNER A., CHINN J., NEUBAUER M.G., PURCHIO A.F.;
RL   DNA CELL BIOL. 10:293-300(1991).
CC   -!- TISSUE SPECIFICITY: TESTIS, SPLEEN, KIDNEY, LUNG, HEART, AND BRAIN
CC       (LOWEST LEVEL IN TESTIS AND HIGHEST IN LUNG).
CC   -!- INDUCTION: BY GROWTH FACTORS.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; M70641; G193314; -.
DR   EMBL; M70642; G193316; -.
DR   EMBL; M80263; G201946; -.
DR   PIR; A53228; A53228.
DR   MGD; MGI:95537; FISP12.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    348       CONNECTIVE TISSUE GROWTH FACTOR.
FT   DOMAIN      100    166       VWFC.
FT   DOMAIN      255    329       CTCK.
FT   DISULFID    255    292       BY SIMILARITY.
FT   DISULFID    272    306       BY SIMILARITY.
FT   DISULFID    283    322       BY SIMILARITY.
FT   DISULFID    286    324       BY SIMILARITY.
FT   DISULFID    291    328       BY SIMILARITY.
FT   CONFLICT    161    161       K -> E (IN REF. 2).
SQ   SEQUENCE   348 AA;  37793 MW;  EAB92BE0 CRC32;


  Query Match             13.7%;  Score 112;  DB 1;  Length 348;
  Best Local Similarity   30.4%;  Pred. No. 3.5e-05;
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|      ||| ||||||||   || |       :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
           |: ::   |     |:|       | :|         :|   |: | | |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   8
IBP5_RAT
ID   IBP5_RAT       STANDARD;      PRT;   271 AA.
AC   P24594;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PRECURSOR (IGFBP-5)
DE   (IBP-5) (IGF-BINDING PROTEIN 5).
GN   IGFBP5 OR IGFBP-5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 20-53.
RC   TISSUE=OVARY;
RX   MEDLINE; 91244847.
RA   SHIMASAKI S., SHIMONAKA M., ZHANG H.-P., LING N.;
RL   J. BIOL. CHEM. 266:10646-10653(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE; 93176146.
RA   ZHU X., LING N., SHIMASAKI S.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 190:1045-1052(1993).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- TISSUE SPECIFICITY: MOSTLY IN KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; M62781; G204746; -.
DR   EMBL; L08275; E73333; -.
DR   PIR; A40403; A40403.
DR   PIR; JC1463; JC1463.
DR   PIR; F40403; F40403.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    271       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN 5.
FT   DOMAIN      214    262       THYROGLOBULIN TYPE I.
SQ   SEQUENCE   271 AA;  30298 MW;  0AA79506 CRC32;


  Query Match             13.5%;  Score 110.5;  DB 1;  Length 271;
  Best Local Similarity   28.7%;  Pred. No. 4.2e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;
```

```
Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAACAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT   9
CE10_CHICK
ID   CE10_CHICK     STANDARD;      PRT;   375 AA.
AC   P19336;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CEF-10 PROTEIN PRECURSOR.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89145206.
RA   SIMMONS D.L., LEVY D.B., YANNONI Y., ERIKSON R.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 86:1178-1182(1989).
CC   -!- FUNCTION: PROBABLE SECRETED REGULATORY PROTEIN.
CC   -!- INDUCTION: BY V-SRC.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; J04496; G211436; -.
DR   PIR; A41428; A41428.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     22
FT   CHAIN        23    375       CEF-10 PROTEIN.
FT   DOMAIN       98    164       VWFC.
FT   DOMAIN      281    355       CTCK.
FT   DISULFID    281    318       BY SIMILARITY.
FT   DISULFID    298    332       BY SIMILARITY.
FT   DISULFID    309    348       BY SIMILARITY.
FT   DISULFID    312    350       BY SIMILARITY.
FT   DISULFID    317    354       BY SIMILARITY.
SQ   SEQUENCE   375 AA;  40651 MW;  68B4BC92 CRC32;


  Query Match             13.5%;  Score 111;  DB 1;  Length 375;
  Best Local Similarity   29.1%;  Pred. No. 4.7e-05;
  Matches   39;  Conservative   19;  Mismatches   40;  Indels   36;  Gaps    9;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRT---VLDDCGCCRVCAAGRG 62
           | ::||  |:|  : :      ||  |: :   ::|:|      | | ||||:|||
Db      10 LAAALLCLARLALGSP------CPAVCQCPA--AAPQCAPGVGLVPDGCGCCKVCAKQLN 61

Qy      63 ETCYRTVSGMDGMKCGP--GLRCQPSNGEDPFGEEFGICKD------CPYG--------T 106
           | | || :      |    || |  : |  | : : |||:       | |          :
Db      62 EDCSRTQP------CDHTKGLEC--NFGASPAATN-GICRAQSEGRPCEYNSKIYQNGES 112

Qy     107 FGMDCRETCNCQSG 120
           |  :|:  | |  |
Db     113 FQPNCKHQCTCIDG 126


RESULT  10
CYR6_HUMAN
ID   CYR6_HUMAN     STANDARD;      PRT;   381 AA.
AC   O00622; O14934;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CYR61 PROTEIN PRECURSOR (GIG1 PROTEIN) (INSULIN-LIKE GROWTH FACTOR-
DE   BINDING PROTEIN 10).
GN   CYR61 OR GIG1 OR IGFBP10.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ALBRECHT C., VON DER KAMMER H., KLAUDINY J., MAYHAUS M., NITSCH R.M.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97280750.
RA   JAY P., BERGE-LEFRANC J.L., MARSOLLIER C., MEJEAN C., TAVIAUX S.,
RA   BERTA P.;
RL   ONCOGENE 14:1753-1757(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   KOLESNIKOVA T.V., LAU L.F.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   BI A.B., YU L.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY ACT AS ONE OF THE MANY GROWTH FACTOR-BINDING
CC       PROTEINS; PROMOTES PROLIFERATION, MIGRATION AND ADHESION (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; Y12084; E311857; -.
DR   EMBL; U62015; G2130527; -.
DR   EMBL; AF003594; G2196782; -.
DR   EMBL; AF031385; G2606094; -.
DR   MIM; 602369; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    381       CYR61 PROTEIN.
FT   DOMAIN       98    164       VWFC.
FT   DOMAIN      286    360       CTCK.
FT   DISULFID    286    323       BY SIMILARITY.
FT   DISULFID    303    337       BY SIMILARITY.
FT   DISULFID    314    353       BY SIMILARITY.
FT   DISULFID    317    355       BY SIMILARITY.
FT   DISULFID    322    359       BY SIMILARITY.
FT   CONFLICT    210    210       L -> I (IN REF. 4).
FT   CONFLICT    220    220       L -> R (IN REF. 4).
SQ   SEQUENCE   381 AA;  42026 MW;  2B091D9E CRC32;


  Query Match             13.5%;  Score 110.5;  DB 1;  Length 381;
  Best Local Similarity   26.1%;  Pred. No. 5.4e-05;
  Matches   36;  Conservative   21;  Mismatches   44;  Indels   37;  Gaps    8;

Qy       2 KSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRT---VLDDCGCCRVCA 58
           ::: |: ||| ::|  :        ||  |  :   :|:|      | | ||||:|||
Db       7 RALALVVTLLHLTRLALS-------TCPAACHCP--LEAPKCAPGVGLVRDGCGCCKVCA 57

Qy      59 AGRGETCYRTVSGMDGMKCGP--GLRCQPSNGEDPFGEEFGICKD------CPYG----- 106
             | | :| :      |    || |  : :  ::    |||:       | |
Db      58 KQLNEDCSKTQP------CDHTKGLECNFGASSTALK---GICRAQSEGRPCEYNSRIYQ 108

Qy     106 ---TFGMDCRETCNCQSG 120
             :|  :|:  | |  |
Db     109 NGESFQPNCKHQCTCIDG 126
```

```
RESULT 11
IBP5_HUMAN
ID   IBP5_HUMAN      STANDARD;      PRT;   272 AA.
AC   P24593;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PRECURSOR (IGFBP-5)
DE   (IBP-5) (IGF-BINDING PROTEIN 5).
GN   IGFBP5 OR IBP5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OSTEOSARCOMA;
RX   MEDLINE; 91207396.
RA   KIEFER M.C., IOH R.S., BAUER D.M., ZAPF J.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 176:219-225(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 91244847.
RA   SHIMASAKI S., SHIMONAKA M., ZHANG H.-P., LING N.;
RL   J. BIOL. CHEM. 266:10646-10653(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94193798.
RA   ALLANDER S.V., LARSSON C., EHRENBORG E., SUWANICHKUL A., WEBER G.,
RA   MORRIS S.L., BAJALICA S., KIEFER M.C., LUTHMAN H., POWELL D.R.;
RL   J. BIOL. CHEM. 269:10891-10898(1994).
RN   [4]
RP   SEQUENCE OF 24-43.
RX   MEDLINE; 91207395.
RA   ANDRESS D.L., BIRNBAUM R.S.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 176:213-218(1991).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- TISSUE SPECIFICITY: OSTEOSARCOMA, AND AT LOWER LEVELS IN LIVER,
CC       KIDNEY, AND BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; M65062; G184820; -.
DR   EMBL; M62782; G184818; -.
DR   EMBL; L27559; G505589; -.
DR   EMBL; L27556; G505589; JOINED.
DR   EMBL; L27557; G505589; JOINED.
DR   EMBL; L27558; G505589; JOINED.
DR   PIR; JH0391; JH0391.
DR   PIR; B40403; B40403.
DR   PIR; PH0143; PH0143.
DR   MIM; 146734; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    272       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN 5.
FT   DOMAIN      215    263       THYROGLOBULIN TYPE I.
SQ   SEQUENCE   272 AA;  30570 MW;  3A6AEEF2 CRC32;


  Query Match             13.4%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 5.4e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |           ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
             || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT 12
CTGF_HUMAN
ID   CTGF_HUMAN      STANDARD;      PRT;   349 AA.
AC   P29279;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CONNECTIVE TISSUE GROWTH FACTOR PRECURSOR.
GN   CTGF.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=UMBILICAL VEIN ENDOTHELIAL CELLS;
RX   MEDLINE; 91373462.
RA   BRADHAM D.M., IGARASHI A., POTTER R.L., GROTENDORST G.R.;
RL   J. CELL BIOL. 114:1285-1294(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=UMBILICAL VEIN ENDOTHELIAL CELLS;
RX   MEDLINE; 93187114.
RA   IGARASHI A., BRADHAM D.M., OKOCHI H., GROTENDORST G.R.;
RL   J. DERMATOL. 19:642-643(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   OEMAR B.S., WERNER A., YANG Z., GARNIER J.M., GENTZ R., LUESCHER T.F.;
RL   SUBMITTED (APR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAJOR CONNECTIVE TISSUE MITOATTRACTANT SECRETED BY
CC       HUMAN VASCULAR ENDOTHELIAL CELLS. THIS IMMEDIATE-EARLY PROTEIN
CC       MAY BIND ONE OF THE PDGF CELL SURFACE RECEPTORS.
CC   -!- SUBUNIT: MONOMER.
CC   -!- ALTERNATIVE PRODUCTS: A SHORTER FORM MAY BE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; M92934; G180924; -.
DR   EMBL; X78947; G474934; -.
DR   PIR; A40551; A40551.
DR   PIR; S44205; S44205.
DR   MIM; 121009; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   GROWTH FACTOR BINDING; SIGNAL; ALTERNATIVE SPLICING.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    349       CONNECTIVE TISSUE GROWTH FACTOR.
FT   DOMAIN      101    167       VWFC.
FT   DOMAIN      256    330       CTCK.
FT   DISULFID    256    293       BY SIMILARITY.
FT   DISULFID    273    307       BY SIMILARITY.
FT   DISULFID    284    323       BY SIMILARITY.
FT   DISULFID    287    325       BY SIMILARITY.
FT   DISULFID    292    329       BY SIMILARITY.
FT   CARBOHYD     28     28       POTENTIAL.
FT   CARBOHYD    225    225       POTENTIAL.
FT   VARSPLIC    172    198       MISSING (IN SHORT FORM).
SQ   SEQUENCE   349 AA;  38069 MW;  C21E9662 CRC32;


  Query Match             13.3%;  Score 109;  DB 1;  Length 349;
  Best Local Similarity   30.4%;  Pred. No. 7.3e-05;
  Matches   34;  Conservative   14;  Mismatches   36;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
```

```
          :|:  |  ::  ::|||      ||| ||||||||  || |       :   | |  ||
Db      28 NCSGPCRCPD-EPAPRCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 80

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
           |: ::   |     |:|       | :|         :|   |:  | |  |
Db      81 FCDFGS---PANRKIGVCTAKDGAPCIFGGTVYRSGESFQSSCKYQCTCLDG 129


RESULT  13
IBP4_SHEEP
ID   IBP4_SHEEP     STANDARD;      PRT;   237 AA.
AC   Q28893;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 (IGFBP-4) (IBP-4)
DE   (IGF-BINDING PROTEIN 4).
GN   IGFBP4.
OS   OVIS ARIES (SHEEP).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 95151165.
RA   CARR J.M., GRANT P.A., FRANCIS G.L., OWENS J.A., WALLACE J.C.,
RA   WALTON P.E.;
RL   J. MOL. ENDOCRINOL. 13:219-236(1994).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; S77394; G944952; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; GLYCOPROTEIN.
FT   DOMAIN       179    228       THYROGLOBULIN TYPE I.
FT   CARBOHYD     104    104       POTENTIAL.
SQ   SEQUENCE   237 AA;  25869 MW;  C1C79FEA CRC32;


  Query Match             13.0%;  Score 107;  DB 1;  Length 237;
  Best Local Similarity   34.8%;  Pred. No. 9.1e-05;
  Matches   24;  Conservative   13;  Mismatches   24;  Indels    8;  Gaps    3;

Qy      25 AVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRGETCYRTVSGMDGMKCGPGL 81
           |: ||   :  : |:::  |   | :  |||| || |:|  |     |:    ||:||
Db       3 AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATCALGKGMPC-----GVYTPDCGSGL 57

Qy      82 RCQPSNGED 90
           ||:|: | :
Db      58 RCHPPRGVE 66


RESULT  14
IBP4_HUMAN
ID   IBP4_HUMAN     STANDARD;      PRT;   258 AA.
AC   P22692;
DT   01-AUG-1991 (REL. 19, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 PRECURSOR (IGFBP-4)
DE   (IBP-4) (IGF-BINDING PROTEIN 4).
GN   IGFBP4 OR IBP4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91186988.
RA   LATOUR D., MOHAN S., LINKHART T.A., BAYLINK D.J., STRONG D.D.;
RL   MOL. ENDOCRINOL. 4:1806-1814(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 91133415.
RA   SHIMASAKI S., UCHIYAMA F., SHIMONAKA M., LING N.;
RL   MOL. ENDOCRINOL. 4:1451-1458(1990).
RN   [3]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 22-41.
RC   TISSUE=OSTEOSARCOMA;
RX   MEDLINE; 91225006.
RA   KIEFER M.C., MASIARZ F.R., BAUER D.M., ZAPF J.;
RL   J. BIOL. CHEM. 266:9043-9049(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   STRONG D.D., MORALES S., LEE K., BOONYARATANAKORNKIT V.,
RA   BAYLINK D.J., MOHAN S.;
RL   SUBMITTED (FEB-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [5]
RP   SEQUENCE OF 22-53.
RC   TISSUE=COLON;
RX   MEDLINE; 91235178.
RA   CULOUSCOU J.-M., SHOYAB M.;
RL   CANCER RES. 51:2813-2819(1991).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- BINDS IGF-II MORE THAN IGF-I.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; M38177; -; NOT_ANNOTATED_CDS.
DR   EMBL; M62403; G184816; -.
DR   EMBL; U20982; G695254; -.
DR   PIR; A36549; A36549.
DR   PIR; B37252; B37252.
DR   PIR; B39842; B39842.
DR   MIM; 146733; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     21
FT   CHAIN       22    258       INSULIN-LIKE GROWTH FACTOR BINDING
FT                               PROTEIN 4.
FT   CARBOHYD    125    125       POTENTIAL.
FT   DOMAIN      200    249       THYROGLOBULIN TYPE I.
FT   CONFLICT     51     51       P -> A (IN REF. 1, 4 AND 5).
FT   CONFLICT    198    198       I -> F (IN REF. 1 AND 4).
SQ   SEQUENCE   258 AA;  27934 MW;  58AC8AC3 CRC32;


  Query Match             13.0%;  Score 106.5;  DB 1;  Length 258;
  Best Local Similarity   31.8%;  Pred. No. 0.00011;
  Matches   28;  Conservative   22;  Mismatches   29;  Indels    9;  Gaps    4;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRG 62
           |:::||::|   :: ::: |: ||   :   : |::  |   | :  |||| || | |
Db       6 LVAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATCALGLG 64

Qy      63 ETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           |     |:   :||:|||| |: | :
Db      65 MPC-----GVYTPRCGSGLRCYPPRGVE 87


RESULT  15
IBP3_PIG
ID   IBP3_PIG       STANDARD;      PRT;   266 AA.
AC   P16611;
DT   01-AUG-1990 (REL. 15, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 3 (IGFBP-3) (IBP-3) (IGF-
```

```
DE   BINDING PROTEIN 3).
GN   IGFBP3.
OS   SUS SCROFA (PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90130475.
RA   SHIMASAKI S., SHIMONAKA M., UI M., INOUYE S., SHIBATA F., LING N.;
RL   J. BIOL. CHEM. 265:2198-2202(1990).
RN   [2]
RP   SEQUENCE OF 1-15.
RX   MEDLINE; 92109718.
RA   COLEMAN M.E., PAN Y.-C.E., ETHERTON T.D.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 181:1131-1136(1991).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND A 85 KD GLYCOPROTEIN (ALS).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- BINDS IGF-II MORE THAN IGF-I.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; J05228; G164501; -.
DR   PIR; A35037; A35037.
DR   PIR; JH0516; JH0516.
DR   HSSP; P17494; 1KST.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; GLYCOPROTEIN.
FT   CARBOHYD     91     91       POTENTIAL.
FT   CARBOHYD    111    111       POTENTIAL.
FT   CARBOHYD    174    174       POTENTIAL.
FT   DOMAIN      211    260       THYROGLOBULIN TYPE I.
FT   CONFLICT      5      6       VG -> A (IN REF. 2).
SQ   SEQUENCE   266 AA;  28910 MW;  53E9830D CRC32;


  Query Match             12.7%;  Score 104.5;  DB 1;  Length 266;
  Best Local Similarity   31.3%;  Pred. No. 0.00018;
  Matches   26;  Conservative   18;  Mismatches   20;  Indels   19;  Gaps    4;

Qy      31 HCDSSECK---------SSPRCKRTVLD-DCGCCRVCAAGRGETCYRTVSGMDGMKCGPG 80
           :|:: : :           ::| |   | :  ||||  ||   |::|     |:   :||:|
Db      12 RCEPCDARALAQCAPPPAAPPCAELVREPGCGCCLTCALREGQAC-----GVYTERCGAG 66

Qy      81 LRCQPSNGE----DPFGEEFGIC 99
           |||||: ||    ::: :  |||
Db      67 LRCQPPPGEPRPLQALLDGRGIC 89


Search completed: May  3, 1999, 16:05:42
Job time: 4001 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 16:59:55 ; Search time 13.05 Seconds
(without alignments)
528.176 Million cell updates/sec

Title: US-09-037-460-2
Perfect score: 918
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....EVVKENAAGSPVMRKWLNPR 184

Scoring table: PAM150

Searched: 116738 seqs, 37460341 residues

Database : PIR_58:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	141.5	15.4	351	2	S20078	NOV protein - chic
2	123.5	13.5	357	2	I38069	gene novH protein
3	115.5	12.6	258	2	A45403	insulin-like growt
4	114.5	12.5	271	2	I48604	insulin-like growt
5	112	12.2	348	2	A40578	beta IG-M2 protein
6	112	12.2	348	2	A53228	fisp-12 protein pr
7	110.5	12.0	271	2	JC1463	insulin-like growt
8	111	12.1	375	2	A41428	CEF-10 protein pre
9	109.5	11.9	272	2	A53748	insulin-like growt
10	109	11.9	349	2	A40551	connective tissue
11	107	11.7	237	2	I47031	insulin-like growt
12	106.5	11.6	258	2	G01662	insulin-like growt
13	106.5	11.6	258	2	B37252	insulin-like growt
14	104.5	11.4	266	2	A35037	insulin-like growt
15	104.5	11.4	271	2	JC4584	insulin-like growt
16	103.5	11.3	254	2	I48599	insulin-like growt
17	103.5	11.3	254	2	JC1464	insulin-like growt
18	103.5	11.3	379	2	A35669	insulin-like growt
19	101.5	11.1	291	1	IOHU3	gene CYR61 protein
20	101.5	11.1	291	1	JN0064	insulin-like growt
21	100.5	10.9	291	2	I48602	insulin-like growt
22	98.5	10.7	292	2	A36748	insulin-like growt
23	98	10.7	251	2	A55035	insulin-like growt
24	103	11.2	1801	1	MMRTS	cysteine-rich prot
25	101.5	11.1	1138	1	S24066	laminin beta-2 cha
26	97	10.6	310	2	A60967	protein-tyrosine k
27	97	10.6	317	2	I46916	insulin-like growt
28	94.5	10.3	259	1	IOHU1	insulin-like growt
29	90.5	9.9	81	2	A45320	insulin-like growt
30	97.5	10.6	1136	1	S57845	transglutaminase s
31	87.5	9.5	111	2	B45403	protein-tyrosine k
32	94.5	10.3	1134	1	JN0711	insulin-like growt
33	94.5	10.3	1786	1	MMMSB1	protein-tyrosine k
34	93.5	10.2	1816	1	S68960	laminin beta-1 cha
35	92.5	10.1	1797	2	A55677	laminin alpha-4 ch
36	92	10.0	1786	1	MMHUB1	laminin beta-2 cha
37	86.5	9.4	328	1	A41927	laminin beta-1 cha
38	85.5	9.3	254	2	I48603	insulin-like growt
39	88	9.6	728	2	I50719	insulin-like growt
40	94	10.2	5147	1	IJFFTM	C-Delta-1 - chicke
41	90	9.8	1639	1	MMFFB2	cadherin-related t
42	84.5	9.2	282	2	S50031	laminin gamma-1 ch
43	84.5	9.2	304	2	A33274	prostacyclin-stimu
44	84.5	9.2	305	2	I48601	insulin-like growt
45	84.5	9.2	305	2	JN0508	insulin-like growt

ALIGNMENTS

RESULT 1
S20078
NOV protein - chicken
C;Species: Gallus gallus (chicken)
C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997
C;Accession: S20078
R;Joliot, V.; Martinerie, C.; Dambrine, G.; Plassiart, G.; Brisac, M.; Crochet, J.; P
Mol. Cell. Biol. 12, 10-21, 1992
A;Title: Proviral rearrangements and overexpression of a new cellular gene (nov) in m
A;Reference number: S20078; MUID:92107157
A;Accession: S20078
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-351 <JOL>
A;Cross-references: EMBL:X59284; NID:g63702; PID:g63703
C;Genetics:
A;Gene: NOV

Query Match 15.4%; Score 141.5; DB 2; Length 351;
Best Local Similarity 31.9%; Pred. No. 2.5e-08;
Matches 45; Conservative 17; Mismatches 38; Indels 41; Gaps 7;

```
Qy       3 SVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAA 59
           :||||  ||: :  |:::        ||: | :     :|||     :||| |||| |||
Db      10 PVLLLLLLLLRPCEVSGR----EAACPRPCGGRCPAEPPRCAPGVPAVLDGCGCCLVCAR 65

Qy      60 GRGETCYRTVSGMDGMKCGPGLRCQPSNG-------EDPFGEEFGICK-----DCPYG-- 106
           |||:                |:| | |: | |       ||  |   |||      :| :
Db      66 QRGES-----------CSPLLPCDESGGLYCDRGPED--GGGAGICMVLEGDNCVFDGM 111

Qy     106 ------TFGMDCRETCNCQSG 120
                 ||   |:  | |: |
Db     112 IYRNGETFQPSCKYQCTCRDG 132
```

RESULT 2
I38069
gene novH protein - human
C;Species: Homo sapiens (man)
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 28-Feb-1997
C;Accession: I38069
R;Martinerie, C.; Huff, V.; Joubert, I.; Badzioch, M.; Saunders, G.; Strong, L.; Perb
Oncogene 9, 2729-2732, 1994
A;Title: Structural analysis of the human nov proto-oncogene and expression in Wilms
A;Reference number: I38069; MUID:94336229
A;Accession: I38069
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-357 <RES>
A;Cross-references: EMBL:X78351; NID:g587422; PID:g825696
C;Genetics:
A;Gene: novH
A;Introns: 28/3; 104/1; 188/1; 259/3

Query Match 13.5%; Score 123.5; DB 2; Length 357;
Best Local Similarity 31.8%; Pred. No. 2.4e-06;
Matches 42; Conservative 25; Mismatches 34; Indels 31; Gaps 9;

```
Qy       7 LTTLLVPAHLVA--AWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGR 61
           || ||:  ||::  | :   : :|| :| :    ::| |    |:||| |:|| |||  |
```

```
Db      18 LTFLLL--HLLGQVAATQRCPPQCPGRCPA----TPPTCAPGVRAVLDGCSCCLVCARQR 71

Qy      62 GETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICK-----DCPYG--------TFG 108
           ||:|    | ::    ::|| |: |   || ::: |||      :| :          |
Db      72 GESC----SDLEPCDESSGLYCDRS--ADP-SNQTGICTAVEGDNCVFDGVIYRSGEKFQ 124

Qy     109 MDCRETCNCQSG 120
             |:  | |: |
Db     125 PSCKFQCTCRDG 136


RESULT   3
A45403
insulin-like growth factor-binding protein 4 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 01-Dec-1995
C;Accession: A45403
R;Moser, D.R.; Lowe Jr., W.L.; Dake, B.L.; Booth, B.A.; Boes, M.; Clemmons, D.R.; Bar, R
Mol. Endocrinol. 6, 1805-1814, 1992
A;Title: Endothelial cells express insulin-like growth factor-binding proteins 2 to 6.
A;Reference number: A45403; MUID:93125553
A;Accession: A45403
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-258 <MOS>
A;Experimental source: pulmonary artery endothelial cells
A;Note: sequence extracted from NCBI backbone (NCBIN:122183, NCBIP:122184)
C;Superfamily: thyroglobulin type I repeat homology
F;174-249/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             12.6%;  Score 115.5;  DB 2;  Length 258;
  Best Local Similarity   32.3%;  Pred. No. 1.4e-05;
  Matches   30;  Conservative   24;  Mismatches   30;  Indels    9;  Gaps    4;

Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVC 57
           | |: |:::||::|   :: ::: |: ||   :   : |:::  |   | : ||||  |
Db       1 MLSLCLMAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATC 59

Qy      58 AAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           | |:|  |     |:   :||:|||| |: | :
Db      60 ALGKGMPC-----GVYTPRCGSGLRCYPPRGVE 87


RESULT   4
I48604
insulin-like growth factor binding protein 5 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 10-Sep-1997
C;Accession: I48604; A48699; A54259
R;Schuller, A.G.; Groffen, C.; van Neck, J.W.; Zwarthoff, E.C.; Drop, S.L.
Mol. Cell. Endocrinol. 104, 57-66, 1994
A;Title: cDNA cloning and mRNA expression of the six mouse insulin-like growth factor bi
A;Reference number: I48600; MUID:95121750
A;Accession: I48604
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-271 <RES>
A;Cross-references: EMBL:X81583; NID:g550384; PID:g550385
R;James, P.L.; Jones, S.B.; Busby Jr., W.H.; Clemmons, D.R.; Rotwein, P.
J. Biol. Chem. 268, 22305-22312, 1993
A;Title: A highly conserved insulin-like growth factor-binding protein (IGFBP-5) is expr
A;Reference number: A48699
A;Accession: A48699
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-271 <JAM>
A;Cross-references: GB:L12447; NID:g425405; PID:g293384
R;Kou, K.; Jenkins, N.A.; Gilbert, D.J.; Copeland, N.G.; Rotwein, P.
Genomics 20, 412-418, 1994
A;Title: Organization, expression, and chromosomal location of the mouse insulin-like gr
A;Reference number: A54259
A;Accession: A54259
A;Molecule type: DNA
A;Residues: 1-111 <KOU>
A;Cross-references: GB:U02023
C;Genetics:
A;Gene: IGFBP-5
C;Superfamily: thyroglobulin type I repeat homology
F;191-262/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             12.5%;  Score 114.5;  DB 2;  Length 271;
  Best Local Similarity   28.7%;  Pred. No. 1.9e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAAYAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT   5
A40578
beta IG-M2 protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Mar-1992 #sequence_revision 06-Mar-1992 #text_change 20-Mar-1998
C;Accession: A40578
R;Brunner, A.; Chinn, J.; Neubauer, M.; Purchio, A.F.
DNA Cell Biol. 10, 293-300, 1991
A;Title: Identification of a gene family regulated by transforming growth factor-beta
A;Reference number: A40578; MUID:91229699
A;Accession: A40578
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-348 <BRU>
A;Cross-references: GB:M80263; NID:g201945; PID:g201946


  Query Match             12.2%;  Score 112;  DB 2;  Length 348;
  Best Local Similarity   30.4%;  Pred. No. 4.4e-05;
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|       ||| ||||||||   || |       :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
           |: ::   |     |:|       | :|         :|   |:  | | |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   6
A53228
fisp-12 protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 10-Sep-1997
C;Accession: A53228
R;Ryseck, R.P.; Macdonald-Bravo, H.; Mattei, M.G.; Bravo, R.
Cell Growth Differ. 2, 225-233, 1991
A;Title: Structure, mapping, and expression of fisp-12, a growth factor-inducible gen
A;Reference number: A53228
A;Accession: A53228
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-348 <RYS>
A;Cross-references: GB:M70641; NID:g193313; PID:g193314
C;Genetics:
A;Gene: fisp-12


  Query Match             12.2%;  Score 112;  DB 2;  Length 348;
  Best Local Similarity   30.4%;  Pred. No. 4.4e-05;
```

```
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|      ||| ||||||||   || |       :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
             |: ::   |     |:|       | :|         :|   |:  | | |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   7
JC1463
insulin-like growth factor-binding protein 5 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 10-Sep-1997
C;Accession: JC1463; A40403; F40403
R;Zhu, X.; Ling, N.; Shimasaki, S.
Biochem. Biophys. Res. Commun. 190, 1045-1052, 1993
A;Title: Cloning of the rat insulin- like growth factor binding protein-5 gene and DNA s
A;Reference number: JC1463
A;Accession: JC1463
A;Molecule type: DNA
A;Residues: 1-271 <ZHU>
A;Cross-references: GB:L08275
R;Shimasaki, S.; Shimonaka, M.; Zhang, H.P.; Ling, N.
J. Biol. Chem. 266, 10646-10653, 1991
A;Title: Identification of five different insulin-like growth factor binding proteins (I
A;Reference number: A40403; MUID:91244847
A;Accession: A40403
A;Molecule type: mRNA
A;Residues: 1-271 <SHI>
A;Cross-references: GB:M62781; NID:g204745; PID:g204746
A;Accession: F40403
A;Status: preliminary
A;Molecule type: protein
A;Residues: 20-25,'X',27-28,'X',30-36,'X',38-43,'X',45-51,'X',53,'XX' <SH2>
C;Genetics:
A;Introns: 112/3; 188/3; 228/3
C;Superfamily: thyroglobulin type I repeat homology
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-271/Product: insulin-like growth factor binding protein 5 #status predicted <MAT>
F;191-262/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             12.0%;  Score 110.5;  DB 2;  Length 271;
  Best Local Similarity   28.7%;  Pred. No. 5.3e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAACAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT   8
A41428
CEF-10 protein precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 10-Sep-1997
C;Accession: A41428
R;Simmons, D.L.; Levy, D.B.; Yannoni, Y.; Erikson, R.L.
Proc. Natl. Acad. Sci. U.S.A. 86, 1178-1182, 1989
A;Title: Identification of a phorbol ester-repressible v-src-inducible gene.
A;Reference number: A41428; MUID:89145206
A;Accession: A41428
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-375 <SIM>
A;Cross-references: GB:J04496; NID:g211435; PID:g211436


  Query Match             12.1%;  Score 111;  DB 2;  Length 375;
  Best Local Similarity   29.1%;  Pred. No. 6e-05;
  Matches   39;  Conservative   19;  Mismatches   40;  Indels   36;  Gaps    9;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRT---VLDDCGCCRVCAAGRG 62
           | ::||  |:|  : :      ||  |: :   ::|:|      | | ||||:|||
Db      10 LAAALLCLARLALGSP------CPAVCQCPA--AAPQCAPGVGLVPDGCGCCKVCAKQLN 61

Qy      63 ETCYRTVSGMDGMKCGP--GLRCQPSNGEDPFGEEFGICKD------CPYG--------T 106
           | | || :      |    || |  : |  | : : |||:       | |           :
Db      62 EDCSRTQP------CDHTKGLEC--NFGASPAATN-GICRAQSEGRPCEYNSKIYQNGES 112

Qy     107 FGMDCRETCNCQSG 120
           |  :|:  | |  |
Db     113 FQPNCKHQCTCIDG 126


RESULT   9
A53748
insulin-like growth factor-binding protein 5 precursor - human
N;Alternate names: IGFBP-5
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 31-Dec-1995 #text_change 20-Mar-1998
C;Accession: A53748; B40403; JH0391; PH0143
R;Allander, S.V.; Larsson, C.; Ehrenborg, E.; Suwanichkul, A.; Weber, G.; Morris, S.L
J. Biol. Chem. 269, 10891-10898, 1994
A;Title: Characterization of the chromosomal gene and promoter for human insulin-like
A;Reference number: A53748
A;Accession: A53748
A;Molecule type: DNA
A;Residues: 1-272 <ALL>
A;Cross-references: GB:L27556; GB:L27557; GB:L27558; GB:L27559; NID:g452057; PID:g505
R;Shimasaki, S.; Shimonaka, M.; Zhang, H.P.; Ling, N.
J. Biol. Chem. 266, 10646-10653, 1991
A;Title: Identification of five different insulin-like growth factor binding proteins
A;Reference number: A40403; MUID:91244847
A;Accession: B40403
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-272 <SHI>
A;Cross-references: GB:M62782; NID:g184817; PID:g184818
A;Experimental source: placenta
R;Kiefer, M.C.; Ioh, R.S.; Bauer, D.M.; Zapf, J.
Biochem. Biophys. Res. Commun. 176, 219-225, 1991
A;Title: Molecular cloning of a new human insulin-like growth factor binding protein.
A;Reference number: JH0391; MUID:91207396
A;Accession: JH0391
A;Molecule type: mRNA
A;Residues: 1-272 <KIE>
A;Cross-references: GB:M65062; NID:g184819; PID:g184820
A;Experimental source: osteosarcoma
R;Andress, D.L.; Birnbaum, R.S.
Biochem. Biophys. Res. Commun. 176, 213-218, 1991
A;Title: A novel human insulin-like growth factor binding protein secreted by osteobl
A;Reference number: PH0143; MUID:91207395
A;Accession: PH0143
A;Molecule type: protein
A;Residues: 24-37,'X',39-41,'X',43 <AND>
A;Experimental source: cell line V-2
C;Comment: Insulin-like growth factors occur in serum and other biological fluids com
C;Genetics:
A;Gene: GDB:IGFBP5
A;Cross-references: GDB:126837
A;Map position: 17q12-17q21
C;Superfamily: thyroglobulin type I repeat homology
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-272/Product: insulin-like growth factor-binding protein 5 #status predicted <IGF
F;192-263/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             11.9%;  Score 109.5;  DB 2;  Length 272;
```

```
  Best Local Similarity   27.7%;  Pred. No. 6.8e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |          ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  10
A40551
connective tissue growth factor - human
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 20-Mar-1998
C;Accession: A40551; S44205
R;Bradham, D.M.; Igarashi, A.; Potter, R.L.; Grotendorst, G.R.
J. Cell Biol. 114, 1285-1294, 1991
A;Title: Connective tissue growth factor: a cysteine-rich mitogen secreted by human vasc
A;Reference number: A40551; MUID:91373462
A;Accession: A40551
A;Molecule type: mRNA
A;Residues: 1-349 <BRA>
A;Cross-references: GB:M92934; GB:M36965; GB:S56201; NID:g180923; PID:g180924
R;Oemar, B.S.; Werner, A.; Yang, Z.; Garnier, J.M.; Gentz, R.; Luescher, T.F.
submitted to the EMBL Data Library, April 1994
A;Description: Differential cloning and expression of human connective tissue growth fac
A;Reference number: S44205
A;Accession: S44205
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-349 <OEM>
A;Cross-references: EMBL:X78947; NID:g474933; PID:g474934


  Query Match             11.9%;  Score 109;  DB 2;  Length 349;
  Best Local Similarity   30.4%;  Pred. No. 9.4e-05;
  Matches   34;  Conservative   14;  Mismatches   36;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           :|:  |  ::  ::|||      ||| ||||||||   || |       :  | |  ||
Db      28 NCSGPCRCPD-EPAPRCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 80

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
           |: ::   |     |:|       | :|         :|   |:  | |  |
Db      81 FCDFGS---PANRKIGVCTAKDGAPCIFGGTVYRSGESFQSSCKYQCTCLDG 129


RESULT  11
I47031
insulin-like growth factor-binding protein-4 - sheep (fragment)
C;Species: Ovis sp. (sheep)
C;Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 10-Oct-1997
C;Accession: I47031
R;Carr, J.M.; Grant, P.A.; Francis, G.L.; Owens, J.A.; Wallace, J.C.; Walton, P.E.
J. Mol. Endocrinol. 13, 219-236, 1994
A;Title: Isolation and characterization of ovine IGFBP-4: protein purification and cDNA
A;Reference number: I47031; MUID:95151165
A;Accession: I47031
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-237 <CAR>
A;Cross-references: GB:S77394; NID:g944951; PID:g944952
C;Superfamily: thyroglobulin type I repeat homology
F;153-228/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             11.7%;  Score 107;  DB 2;  Length 237;
  Best Local Similarity   34.8%;  Pred. No. 0.00012;
  Matches   24;  Conservative   13;  Mismatches   24;  Indels    8;  Gaps    3;

Qy      25 AVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRGETCYRTVSGMDGMKCGPGL 81
           |: ||   :   : |:::  |   | :  ||||  || |:|  |     |:    ||:||
Db       3 AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATCALGKGMPC-----GVYTPDCGSGL 57

Qy      82 RCQPSNGED 90
           ||:|: | :
Db      58 RCHPPRGVE 66


RESULT  12
G01662
insulin-like growth factor binding protein-4 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C;Accession: G01662
R;Strong, D.D.
submitted to the EMBL Data Library, February 1995
A;Reference number: G08078
A;Accession: G01662
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-258 <STR>
A;Cross-references: EMBL:U20982; NID:g695253; PID:g695254
C;Genetics:
A;Gene: GDB:IGFBP4
A;Cross-references: GDB:126811; OMIM:146733
A;Map position: 17q12-17q21
A;Introns: 117/1; 169/3; 214/3


  Query Match             11.6%;  Score 106.5;  DB 2;  Length 258;
  Best Local Similarity   31.8%;  Pred. No. 0.00014;
  Matches   28;  Conservative   22;  Mismatches   29;  Indels    9;  Gaps    4;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRG 62
           |:::||::|   :: ::: |: ||   :   : |:::  |   | :  ||||  || | |
Db       6 LVAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREAGCGCCATCALGLG 64

Qy      63 ETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           |     |:   :||:|||| |: | :
Db      65 MPC-----GVYTPRCGSGLRCYPPRGVE 87


RESULT  13
B37252
insulin-like growth factor-binding protein 4 precursor - human
N;Alternate names: IGFBP-4; inhibitory insulin-like growth factor-binding protein; in
C;Species: Homo sapiens (man)
C;Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 13-Sep-1998
C;Accession: B37252; B39842; A36549; A60712; A54650; A49801; A34419
R;Shimasaki, S.; Uchiyama, F.; Shimonaka, M.; Ling, N.
Mol. Endocrinol. 4, 1451-1458, 1990
A;Title: Molecular cloning of the cDNAs encoding a novel insulin-like growth factor-b
A;Reference number: A37252; MUID:91133415
A;Accession: B37252
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-258 <SHI>
R;Kiefer, M.C.; Masiarz, F.R.; Bauer, D.M.; Zapf, J.
J. Biol. Chem. 266, 9043-9049, 1991
A;Title: Identification and molecular cloning of two new 30-kDa insulin-like growth f
A;Reference number: A39842; MUID:91225006
A;Accession: B39842
A;Molecule type: mRNA
A;Residues: 1-258 <KIE>
A;Cross-references: GB:M62403; NID:g184815; PID:g184816
R;LaTour, D.; Mohan, S.; Linkhart, T.A.; Baylink, D.J.; Strong, D.D.
Mol. Endocrinol. 4, 1806-1814, 1990
A;Title: Inhibitory insulin-like growth factor-binding protein: cloning, complete seq
A;Reference number: A36549; MUID:91186988
A;Accession: A36549
A;Molecule type: mRNA
A;Residues: 1-50,'A',52-197,'F',199-258 <LAT>
```

R;Perkel, V.S.; Mohan, S.; Baylink, D.J.; Linkhart, T.A.
J. Clin. Endocrinol. Metab. 71, 533-535, 1990
A;Title: An inhibitory insulin-like growth factor binding protein (In-IGFBP) from human
A;Reference number: A60712
A;Accession: A60712
A;Molecule type: protein
A;Residues: 22-26,'X',28-29,'X',31-35 <PER>
R;Mohan, S.; Baylink, D.J.
Growth Regul. 1, 110-118, 1991
A;Title: Evidence that the inhibition of TE85 human bone cell proliferation by agents wh
A;Reference number: A54650; MUID:93091814
A;Accession: A54650
A;Molecule type: protein
A;Residues: 22-29,'X',31-42 <MOH>
A;Experimental source: TE85 osteosarcoma cells
A;Note: sequence extracted from NCBI backbone (NCBIP:121076)
R;Culouscou, J.M.; Shoyab, M.
Cancer Res. 51, 2813-2819, 1991
A;Title: Purification of a colon cancer cell growth inhibitor and its identification as
A;Reference number: A49801
A;Accession: A49801
A;Molecule type: protein
A;Residues: 22-53 <CUL>
A;Experimental source: colon adenocarcinoma cells
R;Mohan, S.; Bautista, C.M.; Wergedal, J.; Baylink, D.J.
Proc. Natl. Acad. Sci. U.S.A. 86, 8338-8342, 1989
A;Title: Isolation of an inhibitory insulin-like growth factor (IGF) binding protein fro
A;Reference number: A34419; MUID:90046792
A;Accession: A34419
A;Molecule type: protein
A;Residues: 22-29,'E',31-32,'A',34-36 <MO2>
C;Genetics:
A;Gene: GDB:IGFBP4
A;Cross-references: GDB:126811; OMIM:146733
A;Map position: 17q12-17q21
C;Superfamily: thyroglobulin type I repeat homology
C;Keywords: glycoprotein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-258/Product: insulin-like growth factor-binding protein 4 #status experimental <MAT
F;174-249/Domain: thyroglobulin type I repeat homology <THY1>
F;125/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 11.6%; Score 106.5; DB 2; Length 258;
Best Local Similarity 31.8%; Pred. No. 0.00014;
Matches 28; Conservative 22; Mismatches 29; Indels 9; Gaps 4;

```
Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRG 62
           |:::||::|   :: ::: |: ||   :   : |::: |   | : ||||  || | |
Db       6 LVAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATCALGLG 64

Qy      63 ETCYRTVSGMDGMKCGPGLRCQPSNGED 90
             |     |:   :||:|||| |: | :
Db      65 MPC-----GVYTPRCGSGLRCYPPRGVE 87
```

RESULT 14
A35037
insulin-like growth factor-binding protein 3 homolog - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 20-Jul-1990 #sequence_revision 20-Jul-1990 #text_change 01-Dec-1995
C;Accession: A35037
R;Shimasaki, S.; Shimonaka, M.; Ui, M.; Inouye, S.; Shibata, F.; Ling, N.
J. Biol. Chem. 265, 2198-2202, 1990
A;Title: Structural characterization of a follicle-stimulating hormone action inhibitor
A;Reference number: A35037; MUID:90130475
A;Accession: A35037
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-266 <SHI>
A;Cross-references: GB:J05228
C;Superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I repeat
F;188-260/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 11.4%; Score 104.5; DB 2; Length 266;
Best Local Similarity 31.3%; Pred. No. 0.00024;
Matches 26; Conservative 18; Mismatches 20; Indels 19; Gaps 4;

```
Qy      31 HCDSSECK---------SSPRCKRTVLD-DCGCCRVCAAGRGETCYRTVSGMDGMKCGPG 80
           :|:: : :          ::| |   | :  |||| ||   |::|     |:   :||:|
Db      12 RCEPCDARALAQCAPPPAAPPCAELVREPGCGCCLTCALREGQAC-----GVYTERCGAG 66

Qy      81 LRCQPSNGE----DPFGEEFGIC 99
           |||||: ||    ::: :  |||
Db      67 LRCQPPPGEPRPLQALLDGRGIC 89
```

RESULT 15
JC4584
insulin-like growth factor binding protein-5 - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 10-Apr-1996 #sequence_revision 24-May-1996 #text_change 10-Oct-1997
C;Accession: JC4584; G23734
R;White, M.E.; Diao, R.; Hathaway, M.R.; Mickelson, J.; Dayton, W.R.
Biochem. Biophys. Res. Commun. 218, 248-253, 1996
A;Title: Molecular cloning and sequence analysis of the porcine insulin-like growth f
A;Reference number: JC4584
A;Accession: JC4584
A;Molecule type: mRNA
A;Residues: 1-271 <WHI>
A;Cross-references: GB:U41340; NID:g1173906; PID:g1173907
A;Experimental source: skeletal muscle
R;Shimasaki, S.; Gao, L.; Shimonaka, M.; Ling, N.
Mol. Endocrinol. 5, 938-948, 1991
A;Title: Isolation and molecular cloning of insulin-like growth factor-binding protei
A;Reference number: A23734; MUID:92049376
A;Accession: G23734
A;Molecule type: protein
A;Residues: 20-25,'X',27-28,'X',30-36,'X',38-39 <SHI>
C;Comment: This protein has essential roles in the regulation and coordination of ins
lays a role during myoblast proliferation and differentiation, and is important in th
C;Superfamily: thyroglobulin type I repeat homology
C;Keywords: differentiation; growth factor; skeletal muscle
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-271/Product: insulin-like growth factor binding protein-5 #status experimental <
F;191-262/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 11.4%; Score 104.5; DB 2; Length 271;
Best Local Similarity 27.7%; Pred. No. 0.00024;
Matches 26; Conservative 20; Mismatches 27; Indels 21; Gaps 3;

```
Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :||||::   ||: :::            ||:: : |: : |  : |          ||||
Db       7 LLLLAACAGPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKDPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86
```

Search completed: May 3, 1999, 17:16:49
Job time: 1014 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 22:52:08 ; Search time 49.71 Seconds
(without alignments)
2077.732 Million cell updates/sec

Title: US-09-037-460-1_COPY_61_609
Perfect score: 549
Sequence: 1 AAGAGCGTCTTGCTGCTGAC.....GGAAATGGTTAAATCCACGC 549

Scoring table: IDENTITY_NUC

Searched: 240622 seqs, 94065609 residues

Database : N_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	549	100.0	1271	1	T34991	Vascular IBP-like
2	36.4	6.6	2267	1	T94700	Murine Fisp12 cDNA
3	36.4	6.6	2330	1	Q26422	Gene for beta-IG-M
4	34.4	6.3	230	1	Q36033	Chicken nov gene e
5	35.8	6.5	1975	1	Q36031	Chicken nov coding
6	35.2	6.4	2075	1	T04226	Connective tissue
7	35.2	6.4	2075	1	T51234	Connective tissue
8	35.2	6.4	2075	1	T45360	Human connective t
9	35.2	6.4	2075	1	V38085	Human connective t
10	35.2	6.4	4212	1	V65380	Connective tissue
11	35.2	6.4	4214	1	T58534	Human connective t
12	35.2	6.4	4214	1	T59618	Connective tissue
13	33.8	6.2	1285	1	V29260	Human connective t
14	32.8	6.0	684	1	Q36032	Chicken nov gene f
c 15	31.8	5.8	1614	1	T40206	Tobacco lycopene c
16	31	5.6	1182	1	T99142	S-adenosylmethioni
17	30	5.5	456	1	V46272	Human cortistatin
18	30.4	5.5	993	1	T18036	Transglutaminase c
c 19	30.8	5.6	2319	1	V64070	Arabidopsis thalia
c 20	29.8	5.4	1644	1	V42967	Streptococcus pneu
c 21	29.8	5.4	1644	1	T98604	DNA encoding a S.
c 22	29.4	5.4	960	1	T88154	S. marcesens Ssp-h
c 23	29.4	5.4	960	1	T88155	S. marcesens Ssp-h
c 24	29.4	5.4	1594	1	T35787	Human retinoid X r
c 25	28.2	5.1	295	1	Q75102	Plasmid pOKSC6f co
c 26	28	5.1	269	1	V38066	Human lung tissue
c 27	28.2	5.1	397	1	Q94045	Human ALD gene exo
28	27.8	5.1	266	1	T38906	B2-bradykinin rece
29	27.8	5.1	275	1	T38907	B2-bradykinin rece
30	27.8	5.1	275	1	T38908	B2-bradykinin rece
c 31	28.8	5.2	1401	1	V17721	Human T1 receptor-
c 32	28	5.1	507	1	V38069	Human lung tissue
33	29	5.3	2329	1	V27366	Streptococcus pneu
c 34	28	5.1	519	1	V38070	Human lung tissue
35	27.4	5.0	259	1	T21356	Human gene signatu
c 36	27.2	5.0	201	1	N70194	Signal portion of
c 37	28.8	5.2	2284	1	T31461	Mycobacterial ahpC
38	27	4.9	171	1	T25612	Human gene signatu
39	27.4	5.0	328	1	T20818	Human gene signatu
40	28.4	5.2	1553	1	V43618	Human secreted pro
c 41	27.8	5.1	738	1	V27046	Mouse neurogenin 1
42	28	5.1	1124	1	Q94159	Prostaglandin I2 (
43	28	5.1	1124	1	Q94160	Prostaglandin I2 (
44	27.6	5.0	632	1	Q36034	Chicken nov gene f
45	26.4	4.8	119	1	T25514	Human gene signatu

ALIGNMENTS

```
RESULT   1
T34991
ID   T34991 standard; cDNA; 1271 BP.
AC   T34991;
DT   06-NOV-1996  (first entry)
DE   Vascular IBP-like growth factor cDNA.
KW   Vascular IBF-like growth factor; VIGF;
KW   insulin-like growth factor binding protein; agonist; antagonist;
KW   muscle wastage; osteoporosis; implant fixation; wound healing;
KW   therapy; diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             58. .612
FT                   /*tag= a
FT   signal_peptide  58. .120
FT                   /*tag= b
FT   mat_peptide     121. .609
FT                   /*tag= c
FT   misc_difference 653
FT                   /*tag= d
FT                   /note= "base 653 is given as y in the
FT                   specification"
FT   misc_difference 1096
FT                   /*tag= e
FT                   /note= "base 1096 is given as r in the
FT                   specification"
FT   misc_difference 1212
FT                   /*tag= f
FT                   /note= "base 1212 is given as n in the
FT                   specification"
FT   misc_difference 1225
FT                   /*tag= g
FT                   /note= "base 1225 is given as y in the
FT                   specification"
FT   misc_difference 1255
FT                   /*tag= h
FT                   /note= "base 1255 is given as n in the
FT                   specification"
FT   misc_difference 1259
FT                   /*tag= i
FT                   /note= "base 1259 is given as n in the
FT                   specification"
FT   misc_difference 1261
FT                   /*tag= j
FT                   /note= "base 1261 is given as n in the
FT                   specification"
FT   misc_difference 1265
FT                   /*tag= k
FT                   /note= "base 1265 is given as n in the
FT                   specification"
PN   WO9617931-A1.
PD   13-JUN-1996.
PF   09-DEC-1994; U14388.
PR   09-DEC-1994; WO-U14388.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Hastings GA, Rosen CA;
DR   WPI; 96-287176/29.
DR   P-PSDB; R98994.
PT   Human vascular insulin-like growth factor binding protein-like
PT   growth factor, and its nucleic acid sequence and (ant)agonists -
PT   used, e.g. to treat muscle wasting diseases or aid implant fixation,
PT   or limit excess connective tissue prodn. during wound healing.
PS   Claim 1; Page 42-43; 61pp; English.
CC   A cDNA clone (T34991) codes for human vascular insulin-like
CC   growth factor binding protein-like growth factor (R98994),
CC   or VIGF, a protein structurally related to the IBP and CCN
```

CC families. It was discovered in a cDNA library derived from human
CC umbilical vein endothelial cells. The cDNA can be used for
CC prodn. of recombinant VIGF, e.g. by expression in E. coli, CHO
CC or insect host cells. It is also useful therapeutically e.g.
CC to treat muscle wasting diseases or osteoporosis, or to design
CC probes for the detection of diseases associated with under- or
CC over-expression of VIGF.
SQ Sequence 1271 BP; 360 A; 246 C; 337 G; 320 T;

Query Match 100.0%; Score 549; DB 1; Length 1271;
Best Local Similarity 100.0%; Pred. No. 2.9e-170;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120

Qy      61 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180

Qy     121 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240

Qy     181 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300

Qy     241 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360

Qy     301 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420

Qy     361 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480

Qy     421 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540

Qy     481 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600

Qy     541 AATCCACGC 549
          |||||||||
Db     601 AATCCACGC 609
```

RESULT 2
T94700
ID T94700 standard; cDNA; 2267 BP.
AC T94700;
DT 27-MAR-1998 (first entry)
DE Murine Fispl2 cDNA.
KW Fispl2; cysteine rich protein; mouse; Cyr61;
KW extracellular matrix signalling molecule; cell adhesion;
KW cell migration; cell proliferation; angiogenesis; chrondrogenesis;
KW oncogenesis; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 138. .1284
FT /*tag= a
PN WO9733995-A2.
PD 18-SEP-1997.
PF 14-MAR-1997; U04193.
PR 15-MAR-1996; US-013958.
PA (MUNI-) MUNIN CORP.
PI Lau LF;
DR WPI; 97-470875/43.
DR P-PSDB; W35731.
PT Isolated and purified cysteine rich protein 61, Cyr61 - useful to
PT modulate e.g. haematostasis, induce wound healing, promote organ
PT regeneration etc
PS Example 2; Page 113-114; 133pp; English.
CC This cDNA clone codes for murine Fispl2 (see W35731), an
CC extracellular matrix signalling molecule (ECM) that exhibits
CC structural similarity to Cyr61 (see W35730) and which, like
CC Cyr61, influences cell adhesion, proliferation and migration.
CC The human orthologue of Fispl2 is connective tissue growth
CC factor. Fispl2 polynucleotides can be used for the production
CC of Fispl2 polypeptides by recombinant methods. Polypeptide
CC compositions are provided that comprise mammalian ECM signalling
CC molecules, peptide fragments, inhibitory peptides capable of
CC interacting with receptors for ECM signalling molecules, and
CC antibody products. Further provided are methods for using
CC mammalian ECM signalling molecules to screen for, and/or modulate
CC disorders associated with angiogenesis, chondrogenesis and
CC oncogenesis; ex vivo methods for using ECM signalling molecules
CC to prepare blood products are also provided.
SQ Sequence 2267 BP; 570 A; 568 C; 582 G; 547 T;

Query Match 6.6%; Score 36.4; DB 1; Length 2267;
Best Local Similarity 66.7%; Pred. No. 0.037;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
Qy     114 CAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGC 173
           | ||||  | ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db     260 CTGCCCCGCCGGCGTGAGCCTGGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCAA 319

Qy     174 AGGGCGGGGAGAAACTTG 191
             || |||||||   ||
Db     320 GCAGCTGGGAGAACTGTG 337
```

RESULT 3
Q26422
ID Q26422 standard; DNA; 2330 BP.
AC Q26422;
DT 18-JAN-1993 (first entry)
DE Gene for beta-IG-M2.
KW Transforming growth factor beta; induced; CEF-10; v-src; chicken;
KW embryo; fibroblasts; TGF-beta; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 204. .1247
FT /*tag= a
PN EP-495674-A.
PD 22-JUL-1992.
PF 17-JAN-1992; 300429.
PR 18-JAN-1991; US-642991.
PR 10-JAN-1992; US-816270.
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Brunner AM, Chinn J, Neubauer MG, Purchio AF;
DR WPI; 92-243508/30.
DR P-PSDB; R25565.
PT TGF-beta induced gene family - encodes proteins involved in
PT growth and differentiation effects of TGF-beta-1
PS Claim 7; Fig 2; 35pp; English.
CC The DNA encoding mouse beta-IG-M2 was obtd. from AKR-2B mouse cells
CC induced with TGF-beta1 and cyclohexamide. Poly RNA extracted from
CC these cells was used to create a cDNA library which was screened
CC using two probes. The probes were prepd. from untreated AKR-2B mRNA
CC and AKR-2B mRNA from cells treated with cyclohexamide and TGF-beta1.
CC Hybridising colonies were isolated and two clones (beta-IG-M1 and
CC beta-IG-M2) were then sequenced. The DNA encodes proteins that
CC have a 80 and 50 percent homology respectively with the CEF-10
CC protein induced by v-src in chicken embryo fibroblasts. The
CC proteins encoded by TGF-beta induced genes are likely to be

```
CC   involved in mediation of the biological effects of TGF-beta
CC   relating to cell growth and differentiation.
CC   See also Q26421.
SQ   Sequence   2330 BP;    589 A;    598 C;    595 G;    548 T;


  Query Match             6.6%;  Score 36.4;  DB 1;  Length 2330;
  Best Local Similarity   66.7%;  Pred. No. 0.037;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy     114 CAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGC 173
           | ||||   | ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db     326 CTGCCCCGCCGGCGTGAGCCTGGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCAA 385

Qy     174 AGGGCGGGGAGAAACTTG 191
               || |||||||   ||
Db     386 GCAGCTGGGAGAACTGTG 403


RESULT   4
Q36033
ID   Q36033 standard; cDNA; 230 BP.
AC   Q36033;
DT   24-MAY-1993  (first entry)
DE   Chicken nov gene exon 2 fragment III.
KW   avian nephroblastoma; avian myeloblastoma virus;
KW   stringent hybridisation; ss.
OS   Gallus domesticus.
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
DR   P-PSDB; R31600.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 2; Page 27; 67pp; French.
CC   The chicken nov gene is stimulated in avian nephroblastoma induced
CC   by avian myeloblastoma virus but not in normal adult kidney. A
CC   1975bp cDNA sequence was isolated from a gene bank prepared from
CC   chicken embryonic fibroblasts screened with a tumour-derived probe.
CC   Fragment III is derived from the 2nd. exon of the nov gene;
CC   nucleotide sequences which hybridise to Fragment III under stringent
CC   conditions (i.e. 50% formamide, 5 x SCC) are claimed. The claimed
CC   sequences preferably encode a protein with at least 70% homology to
CC   R31600 which is encoded by Fragment III.
SQ   Sequence   230 BP;    18 A;    89 C;    97 G;    26 T;


  Query Match             6.3%;  Score 34.4;  DB 1;  Length 230;
  Best Local Similarity   63.1%;  Pred. No. 0.053;
  Matches   53;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy     113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
           || || | |   |   |    | ||||| |||| ||| ||||||||||  ||||||||
Db      67 GCTGCGCCCCGGGAGTGCCCGCCGTGCTGGACGGCTGCGGCTGCTGCCTGGTGTGCGCCC 126

Qy     173 CAGGGCGGGGAGAAACTTGCTACC 196
               ||| || || |  |||| ||
Db     127 GGCAGCGCGGCGAGAGCTGCTCCC 150


RESULT   5
Q36031
ID   Q36031 standard; cDNA; 1975 BP.
AC   Q36031;
DT   24-MAY-1993  (first entry)
DE   Chicken nov coding sequence.
KW   avian nephroblastoma; avian myeloblastoma virus; ss.
OS   Gallus domesticus.

FH   Key             Location/Qualifiers
FT   cds             24. .1079
FT                   /*tag= a
FT   polya_signal    1814. .1818
FT                   /*tag= b
FT   polya_signal    1932. .1937
FT                   /*tag= c
FT   signal_peptide  24. .95
FT                   /*tag= d
FT   exon            1. .92
FT                   /*tag= e
FT                   /number= 1
FT   exon            93. .321
FT                   /*tag= f
FT                   /number= 2
FT   exon            322. .573
FT                   /*tag= g
FT                   /number= 3
FT   exon            574. .782
FT                   /*tag= h
FT                   /number= 4
FT   exon            783. .1975
FT                   /*tag= i
FT                   /number= 5
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
DR   P-PSDB; R31599.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 1; Fig 1; 67pp; French.
CC   The chicken nov gene is stimulated in avian nephroblastoma induced
CC   by avian myeloblastoma virus but not in normal adult kidney. The
CC   1975bp cDNA sequence was isolated from a gene bank prepared from
CC   chicken embryonic fibroblasts screened with a tumour-derived probe.
CC   The cDNA was cloned in Bluescript KS+ to form plasmid pC1K which was
CC   labelled and used to screen a human placental DNA library. A nov-
CC   analogous gene was isolated from the human library. Nucleotide
CC   sequences which hybridise to the chicken nov gene under stringent
CC   conditions (i.e. 50% formamide, 5 x SCC) are claimed.
SQ   Sequence   1975 BP;    522 A;    451 C;    486 G;    516 T;


  Query Match             6.5%;  Score 35.8;  DB 1;  Length 1975;
  Best Local Similarity   52.9%;  Pred. No. 0.054;
  Matches  101;  Conservative    0;  Mismatches   87;  Indels    3;  Gaps    1;

Qy       6 CGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAATAA 65
           |||| |||||||| |  |||||||| | |   |   |  ||||  ||   ||
Db      53 CGTCCTGCTGCTGCTCCTGCTCCTCCTCCGGCCGTGCGAGGTGAGCGGGCGGGAGGCGGC 112

Qy      66 TTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCGCTG 125
             |   |   | ||||      | |||   |  |   |||        || || | |   |
Db     113 GTGCCCCCGGCCCTGCGGCGGGCGCTG---CCCCGCGGAGCCGCCGCGCTGCGCCCCGGG 169

Qy     126 CAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGGAGA 185
             |    | ||||| |||| ||| ||||||||||  ||||||||      ||| || ||
Db     170 AGTGCCCGCCGTGCTGGACGGCTGCGGCTGCTGCCTGGTGTGCGCCCGGCAGCGCGGCGA 229

Qy     186 AACTTGCTACC 196
             |  |||| ||
Db     230 GAGCTGCTCCC 240


RESULT   6
T04226
ID   T04226 standard; cDNA; 2075 BP.
AC   T04226;
```

```
DT   12-JUN-1996  (first entry)
DE   Connective tissue growth factor cDNA.
KW   Connective tissue growth factor; CTGF; wound healing; vulnerary;
KW   cell proliferation; cancer; fibrosis; atherosclerosis; diagnosis
KW   therapy; mitogen; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             130. .1179
FT                   /*tag=  a
PN   US5408040-A.
PD   18-APR-1995.
PF   30-AUG-1991; 752427.
PR   30-AUG-1991; US-752427.
PR   14-DEC-1993; US-167628.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM,  Grotendorst GR;
DR   WPI; 95-161147/21.
DR   P-PSDB; R79964.
PT   New connective tissue growth factor - used to develop prods. for
PT   wound healing and for diagnosis and therapy of cell proliferative
PT   disorders.
PS   Example 5; Column 15-18; 12pp; English.
CC   A cDNA clone (T04226), designated DB60R32, codes for human
CC   connective tissue growth factor (CTGF) (R79964), a protein that
CC   has mitogenic and chemotactic activity for connective tissue
CC   cells and which binds to the platelet-derived growth factor
CC   (PDGR) receptor.  The cDNA was obtd. from a library of human
CC   umbilical vein endothelial cell cDNA in vector lambda gt11 by
CC   screening with an anti-PDGF antibody.  The cDNA is useful for
CC   prodn. of recombinant CTGF or in the design of antisense or
CC   ribozyme constructs used to treat disorders associated with
CC   overgrowth of tissue cells, such as cancer, fibrotic diseases
CC   and atherosclerosis.
SQ   Sequence   2075 BP;    491 A;    558 C;    546 G;    480 T;

  Query Match             6.4%;  Score 35.2;  DB 1;  Length 2075;
  Best Local Similarity   65.0%;  Pred. No. 0.087;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
           || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db     254 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 313

Qy     173 CAGGGCGGGGAGAAACTTGC 192
             || ||| ||    |||
Db     314 AGCAGCTGGGCGAGCTGTGC 333


RESULT   7
T51234
ID   T51234 standard; cDNA; 2075 BP.
AC   T51234;
DT   18-MAR-1997  (first entry)
DE   Connective tissue growth factor coding sequence.
KW   Connective tissue growth factor; CTGF; human; connective tissue cell;
KW   proliferative disease; platelet-derived growth factor; PDGF; development;
KW   tissue growth; repair; umbilical vein endothelial cell; HUVE cell;
KW   antibody; wound healing; cancer; fibrotic disease; atherosclerosis;
KW   inhibitor; protease degradation; growth factor; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             130. .1179
FT                   /*tag=  a
FT                   /product= connective tissue growth factor
PN   US5585270-A.
PD   17-DEC-1996.
PF   30-AUG-1991; 752427.
PR   30-AUG-1991; US-752427.
PR   14-DEC-1993; US-167628.
PR   10-FEB-1995; US-386680.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM,  Grotendorst GR;
DR   WPI; 97-051180/05.
DR   P-PSDB; W11302.
PT   New nucleic acid encoding connective tissue growth factor - useful
PT   for accelerating wound healing, also for diagnosis and treatment of
PT   proliferative disease
PS   Claim 10; Column 15-18; 11pp; English.
CC   This sequence represents the cDNA encoding the human connective tissue
CC   growth factor (CTGF).  CTGF is related immunologically and biologically
CC   to platelet-derived growth factor (PDGF), but is encoded by an unrelated
CC   gene.  CTGF is thought to play a significant role in the normal
CC   development, growth, and repair of human tissue, similarly to PDGF.  This
CC   sequence was isolated by screening a cDNA library from human umbilical
CC   vein endothelial (HUVE) cells with anti-PDGF antibodies.  CTGF can be
CC   used to accelerate wound healing.  Also, elevated levels of CTGF may be
CC   diagnostic of proliferative diseases involving outgrowth of connective
CC   tissue cells, such as cancer, fibrotic disease and atherosclerosis.  All
CC   of these diseases can be treated with reagents reactive with CTGF, such
CC   as antibodies (which can also serve as assay reagents).  Antisense
CC   nucleic acids, and ribozymes could also be used to inhibit CTGF
CC   production.  The advantage with using CTGF is that it is more stable, and
CC   less susceptible to protease degradation than PDGF, and other growth
CC   factors involved in wound healing.  This is believed to be due to the
CC   high Cys content.
SQ   Sequence   2075 BP;    491 A;    558 C;    546 G;    480 T;

  Query Match             6.4%;  Score 35.2;  DB 1;  Length 2075;
  Best Local Similarity   65.0%;  Pred. No. 0.087;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
           || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db     254 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 313

Qy     173 CAGGGCGGGGAGAAACTTGC 192
             || ||| ||    |||
Db     314 AGCAGCTGGGCGAGCTGTGC 333


RESULT   8
T45360
ID   T45360 standard; cDNA; 2075 BP.
AC   T45360;
DT   26-APR-1997  (first entry)
DE   Human connective tissue growth factor cDNA.
KW   Connective tissue growth factor; CTGF; mitogen; cell proliferation;
KW   wound healing; cancer; tumour; fibrosis; glaucoma; atherosclerosis;
KW   diagnosis; therapy; antisense; triple helix; ribozyme; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             130. .1179
FT                   /*tag=  a
PN   WO9638172-A1.
PD   05-DEC-1996.
PF   31-MAY-1996; U08140.
PR   31-MAY-1996; WO-U08140.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM,  Grotendorst GR;
DR   WPI; 97-042659/04.
DR   P-PSDB; W09089.
PT   Connective tissue growth factor coding sequence and protein  - used
PT   in the treatment of proliferative disorders and to accelerate wound
PT   healing
PS   Claim 20; Page 50-52; 76pp; English.
CC   A cDNA clone (T45360) codes for novel human connective tissue growth
CC   factor (CTGF) (W09089), a PDGF-immunorelated protein that may play a
CC   significant role in the normal development, growth and repair of
CC   human tissue and which probably functions as a growth factor in
CC   wound healing.  CTGF may be involved in diseases in which there is
CC   an overgrowth of connective tissue cells, such as cancer, tumour
CC   formation and growth, fibrotic diseases (e.g. pulmonary fibrosis,
CC   kidney fibrosis, glaucoma) and atherosclerosis.  The cDNA clone was
CC   isolated from a HUVEC cDNA library using anti-PDGF antibody.  CTGF
```

```
CC   nucleic acids can be used for recombinant prodn. of CTGF and as
CC   probes to detect CTGF mRNA. CTGF genomic DNA (T58534) has also
CC   been isolated. A proliferative disorder may be treated using e.g.
CC   a CTGF antisense, ribozyme or triplex agent.
SQ   Sequence  2075 BP;  491 A;  558 C;  546 G;  480 T;


  Query Match              6.4%;  Score 35.2;  DB 1;  Length 2075;
  Best Local Similarity   65.0%;  Pred. No. 0.087;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
          || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    254 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 313

Qy    173 CAGGGCGGGGAGAAACTTGC 192
              || ||| ||    |||
Db    314 AGCAGCTGGGCGAGCTGTGC 333


RESULT   9
V38085
ID   V38085 standard; cDNA; 2075 BP.
AC   V38085;
DT   15-SEP-1998  (first entry)
DE   Human connective tissue growth factor encoding cDNA.
KW   Human; connective tissue growth factor; CTGF; PDGF; diagnosis; cancer;
KW   platelet derived growth factor; ameliorating cell proliferative disorder;
KW   atherosclerosis; fibrotic disease; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             130. .1179
FT                   /*tag= a
FT                   /product= "connective tissue growth factor"
PN   US5783187-A.
PD   21-JUL-1998.
PF   11-SEP-1996; 712302.
PR   30-AUG-1991; US-752427.
PR   14-DEC-1993; US-167628.
PR   11-SEP-1996; US-712302.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM, Grotendorst GR;
DR   WPI; 98-426958/36.
DR   P-PSDB; W62084.
PT   Ameliorating cell proliferative disorder associated with connective
PT   tissue growth factor - comprises the administration of an antibody
PT   which binds to connective tissue growth factor and not to
PT   platelet-derived growth factor
PS   Example 6; Column 15-18; 11pp; English.
CC   A method has been developed for ameliorating a cell proliferative
CC   disorder associated with connective tissue growth factor (CTGF). The
CC   method comprises the administration of an antibody or its fragment that
CC   binds to CTGF and not to platelet-derived growth factor (PDGF), to the
CC   site of the disorder. CTGF is related immunologically and biologically
CC   to PDGF. The present sequence encodes CTGF. The method is used to treat
CC   conditions involving the overgrowth of connective tissue cells such
CC   as cancer, atherosclerosis and other fibrotic diseases.
SQ   Sequence  2075 BP;  491 A;  558 C;  546 G;  480 T;


  Query Match              6.4%;  Score 35.2;  DB 1;  Length 2075;
  Best Local Similarity   65.0%;  Pred. No. 0.087;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
          || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    254 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 313

Qy    173 CAGGGCGGGGAGAAACTTGC 192
              || ||| ||    |||
Db    314 AGCAGCTGGGCGAGCTGTGC 333


RESULT  10
V65380
ID   V65380 standard; DNA; 4212 BP.
AC   V65380;
DT   25-JAN-1999  (first entry)
DE   Connective tissue growth factor (CTGF) genomic DNA sequence.
KW   CTGF; connective tissue growth factor; bone formation; tissue; arthritis;
KW   wound healing; cartilage formation; osteoporosis; osteoarthritis; burn;
KW   osteochondrytis; skeletal disorder; hypertrophic scar; protease; PDGF;
KW   degradation; vascular hypertrophy; platelet derived growth factor; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   5'UTR           1. .1022
FT                   /*tag= a
FT   CDS             1023. .2928
FT                   /*tag= b
FT                   /product= "CTGF"
FT                   /note= "contains introns"
FT   sig_peptide     1023. .1085
FT                   /*tag= c
FT   mat_peptide     1086. .2925
FT                   /*tag= d
FT   exon            1023. .1088
FT                   /*tag= e
FT                   /number= 1
FT   intron          1089. .1202
FT                   /*tag= f
FT                   /number= 1
FT   exon            1203. .1425
FT                   /*tag= g
FT                   /number= 2
FT   intron          1426. .1651
FT                   /*tag= h
FT                   /number= 2
FT   exon            1652. .1902
FT                   /*tag= i
FT                   /number= 3
FT   intron          1903. .2032
FT                   /*tag= j
FT                   /number= 3
FT   exon            2033. .2244
FT                   /*tag= k
FT                   /number= 4
FT   intron          2245. .2631
FT                   /*tag= l
FT                   /number= 4
FT   exon            2632. .2925
FT                   /*tag= m
FT                   /number= 5
FT   3'UTR           2928. .4212
FT                   /*tag= n
FT   polyA_signal    3984. .3989
FT                   /*tag= o
PN   US5837258-A.
PD   17-NOV-1998.
PF   31-MAY-1996; 656393.
PR   31-MAY-1996; US-656393.
PR   30-AUG-1991; US-752427.
PR   14-DEC-1993; US-167628.
PR   10-FEB-1995; US-386680.
PR   02-JUN-1995; US-459717.
PA   (UYMI-) UNIV MIAMI.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Grotendorst GR;
DR   WPI; 99-023382/02.
DR   P-PSDB; W81425#.
PT   Use of connective tissue growth factor - for inducing bone, tissue
PT   or cartilage formation in a patient or for inducing wound healing
PS   Disclosure; Fig 1C; 30pp; English.
CC   This represents the genomic DNA sequence of connective tissue growth
CC   factor (CTGF). This can be used in the method of the invention for
CC   inducing bone or tissue formation that comprises administration to a
CC   patient, a composition comprising CTGF and a carrier. CTGF can also be
```

```
CC   used in a method for inducing wound healing. The methods can be used for
CC   inducing bone, tissue or cartilage formation in disorders such as
CC   osteoporosis, osteoarthritis or osteochondrytis, arthritis, skeletal
CC   disorders, hypertrophic scars, burns, vascular hypertrophy, or in wound
CC   healing. The CTGF and functional fragments are more stable and less
CC   susceptible to protease degradation than platelet derived growth factor
CC   (PDGF) and other growth factors known to be involved in wound healing.
SQ   Sequence   4212 BP;    1038 A;    1080 C;    1008 G;    1086 T;


  Query Match             6.4%;  Score 35.2;  DB 1;  Length 4212;
  Best Local Similarity   65.0%;  Pred. No. 0.12;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
           || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    1261 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 1320

Qy     173 CAGGGCGGGGAGAAACTTGC 192
             || ||| ||    |||
Db    1321 AGCAGCTGGGCGAGCTGTGC 1340


RESULT  11
T58534
ID   T58534 standard; cDNA; 4214 BP.
AC   T58534;
DT   26-APR-1997  (first entry)
DE   Human connective tissue growth factor gene.
KW   Connective tissue growth factor; CTGF; mitogen; cell proliferation;
KW   wound healing; cancer; tumour; fibrosis; glaucoma; atherosclerosis;
KW   diagnosis; therapy; antisense; triple helix; ribozyme; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   promoter        1. .1024
FT                   /*tag= a
FT                   /note= "promoter region (-823 to +200 relative
FT                   to the transcription start site)"
FT   misc_feature    173. .179
FT                   /*tag= b
FT                   /label= AP-1
FT   misc_feature    443. .452
FT                   /*tag= c
FT                   /label= CArG_box
FT   misc_feature    533. .539
FT                   /*tag= d
FT                   /label= AP-1
FT   misc_feature    630. .642
FT                   /*tag= e
FT                   /label= NF-1_like
FT   misc_feature    660. .705
FT                   /*tag= f
FT                   /note= "transforming growth factor-beta response
FT                   element (Claim 31)
FT   misc_feature    667. .679
FT                   /*tag= g
FT                   /label= T-beta-RE
FT                   /note= "transforming growth factor-beta response
FT                   element (Claim 32)"
FT   misc_feature    696. .705
FT                   /*tag= h
FT                   /label= TIE-like
FT                   /note= "transforming growth factor-beta inhibito"
FT   misc_feature    766. .770
FT                   /*tag= i
FT                   /label= CATbox
FT   misc_feature    782. .787
FT                   /*tag= j
FT                   /label= SP-1
FT   tata_signal     794. .800
FT                   /*tag= k
FT   misc_feature    809. .818
FT                   /*tag= l
FT                   /label= SP-1
FT   misc_feature    824
FT                   /*tag= m
FT                   /function= transcription start site
FT   exon            1. .1090
FT                   /*tag= n
FT                   /codon_start= 1025. .1027
FT   intron          1091. .1204
FT                   /*tag= o
FT   exon            1205. .1427
FT                   /*tag= p
FT   intron          1428. .1653
FT                   /*tag= q
FT   exon            1654. .1905
FT                   /*tag= r
FT   intron          1906. .2034
FT                   /*tag= s
FT   exon            2035. .2306
FT                   /*tag= t
FT   intron          2307. .2633
FT                   /*tag= u
FT   exon            2634. .4214
FT                   /*tag= v
FT                   /note= "in-frame stop codon at bases 2928-2930"
FT   polya_signal    3986. .3991
FT                   /*tag= w
FT   polya_signal    4138. .4143
FT                   /*tag= x
PN   WO9638172-A1.
PD   05-DEC-1996.
PF   31-MAY-1996; U08140.
PR   31-MAY-1996; WO-U08140.
PA   (UYSF-) UNIV SOUTH FLORIDA.
PI   Bradham DM,  Grotendorst GR;
DR   WPI; 97-042659/04.
DR   P-PSDB; W09089.
PT   Connective tissue growth factor coding sequence and protein  - used
PT   in the treatment of proliferative disorders and to accelerate wound
PT   healing
PS   Claim 22; Page Fig 1C; 76pp; English.
CC   The human CTGF gene (T58534) codes for novel human connective
CC   tissue growth factor (CTGF) (W09089), a PDGF-immunorelated protein
CC   that may play a significant role in the normal development, growth
CC   and repair of human tissue and which probably functions as a growth
CC   factor in wound healing.  The gene was isolated from a human
CC   library using a partial gene fragment as probe.  A CTGF cDNA clone
CC   (see also T45360) has also been obtd.  The 5' nontranslated region
CC   of the CTGF gene functions as a transforming growth factor-beta
CC   (TGF-beta) inducible promoter.  The TGF-beta regulatory element
CC   was identified (see also T49433) and can be used to identify cpds.
CC   the affect CTGF expression.
SQ   Sequence   4214 BP;    1039 A;    1079 C;    1006 G;    1090 T;


  Query Match             6.4%;  Score 35.2;  DB 1;  Length 4214;
  Best Local Similarity   65.0%;  Pred. No. 0.12;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
           || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    1263 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 1322

Qy     173 CAGGGCGGGGAGAAACTTGC 192
             || ||| ||    |||
Db    1323 AGCAGCTGGGCGAGCTGTGC 1342


RESULT  12
T59618
ID   T59618 standard; DNA; 4214 BP.
AC   T59618;
DT   31-MAY-1997  (first entry)
DE   Connective tissue growth factor gene.
```

```
KW   Connective tissue growth factor; CTGF; bone; cartilage; vulnerary;
KW   wound healing; osteoporosis; osteoarthritis; osteochondrytis; ss.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   promoter        1. .823
FT                   /*tag= a
FT                   /note= "CTGF gene promoter region"
FT   misc_feature    173. .179
FT                   /*tag= b
FT                   /label= AP-1
FT   misc_feature    443. .452
FT                   /*tag= c
FT                   /label= CArG_box
FT   misc_feature    630. .642
FT                   /*tag= d
FT                   /label= NF-1_like
FT   misc_feature    696. .705
FT                   /*tag= e
FT                   /label= TIE_like
FT   misc_feature    766. .771
FT                   /*tag= f
FT                   /label= CATbox
FT   misc_feature    782. .787
FT                   /*tag= g
FT                   /label= SP-1
FT   tata_signal     794. .800
FT                   /*tag= h
FT                   /label= TATAbox
FT   misc_feature    809. .818
FT                   /*tag= i
FT                   /label= SP-1
FT   exon            1. .1090
FT                   /*tag= j
FT                   /codon_start= 1025. .1028
FT   intron          1091. .1204
FT                   /*tag= k
FT                   /label= Intron-1
FT   exon            1205. .1427
FT                   /*tag= l
FT   intron          1428. .1652
FT                   /*tag= m
FT                   /label= Intron-2
FT   exon            1653. .1905
FT                   /*tag= n
FT   intron          1906. .2034
FT                   /*tag= o
FT                   /label= Intron-3
FT   exon            2035. .2246
FT                   /*tag= p
FT   intron          2247. .2633
FT                   /*tag= q
FT                   /label= Intron-4
FT   exon            2634. .4214
FT                   /*tag= r
FT                   /note= "in-frame stop codon at position 2928-2930"
FT   polya_signal    3986. .1991
FT                   /*tag= s
FT   polya_signal    4138. .4143
FT                   /*tag= t
PN   WO9638168-A1.
PD   05-DEC-1996.
PF   31-MAY-1996; U08210.
PR   31-MAY-1996; WO-U08140.
PA   (GROT/) GROTENDORST G R.
PI   Grotendorst GR;
DR   WPI; 97-042658/04.
DR   P-PSDB; W12694.
PT   Connective Tissue Growth Factor composition - for inducing bone,
PT   tissue and cartilage formation and wound healing
PS   Disclosure; Fig 1C1-3; 60pp; English.
CC   The CTGF gene (T59618) codes for connective tissue growth factor
CC   (W12694), a cysteine-rich mitogenic protein which is selectively
CC   induced in fibroblasts after activation with transforming growth
CC   factor beta. The gene may be incorporated into a vector and
CC   utilised in the prodn. of recombinant CTGF in prokaryotic or
CC   eukaryotic host cells, or used to design probes and primers. The
CC   promoter region of the gene can be used in methods for the
CC   identification of cpds. which enhance the activity of CTGF, for
CC   use in inducing bone, tissue and cartilage and for wound
CC   healing.
SQ   Sequence   4214 BP;    1039 A;    1078 C;    1009 G;    1088 T;
```

```
  Query Match             6.4%;  Score 35.2;  DB 1;  Length 4214;
  Best Local Similarity   65.0%;  Pred. No. 0.12;
  Matches   52;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
           || |||||    ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    1263 GCTGCCCGGCGGGCGTGAGCCTCGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCA 1322

Qy     173 CAGGGCGGGGAGAAACTTGC 192
             || ||| ||   |||
Db    1323 AGCAGCTGGGCGAGCTGTGC 1342
```

```
RESULT  13
V29260
ID   V29260 standard; DNA; 1285 BP.
AC   V29260;
DT   14-SEP-1998  (first entry)
DE   Human connective tissue growth factor-3 gene.
KW   Connective tissue growth factor-3; CTGF-3; human; cancer;
KW   arthritis; fibrosis; osteoporosis; diagnosis; therapy; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             9. .761
FT                   /*tag= a
FT   sig_peptide     9. .65
FT                   /*tag= b
FT   mat_peptide     66. .758
FT                   /*tag= c
PN   WO9821236-A1.
PD   22-MAY-1998.
PF   08-NOV-1996; U17856.
PR   08-NOV-1996; WO-U17856.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Chopra A, Ebner R, Ruben SM;
DR   WPI; 98-297864/26.
DR   P-PSDB; W37946.
PT   Novel human connective tissue growth factor 3 gene - useful for the
PT   diagnosis and treatment of e.g. cancer, arthritis, fibrosis,
PT   osteoporosis
PS   Claim 2; Fig 1A-B; 87pp; English.
CC   This nucleotide sequence codes for human connective tissue growth
CC   factor-3 (CTGF-3) protein (see W37946), a novel member of the
CC   growth factor superfamily.  It was discovered in a cDNA library
CC   derived from human osteoblasts.  The gene has also been identified
CC   in cDNA libraries from ovary, testis, heart, lung, skeletal muscle,
CC   adrenal medulla, adrenal cortex, thymus, prostate, small intestine
CC   and colon.  A cDNA clone is deposited as ATCC 97756.  Also provided
CC   are vectors, host cells and recombinant methods for producing
CC   CTGF-3 polypeptides.  CTGF-3 nucleic acid sequences or their
CC   fragments, e.g. primers or probes, can be used to diagnose diseases
CC   where CTGF-3 expression is enhanced, e.g. cancer, arthritis,
CC   fibrosis or atherosclerosis, or diseases where expression is
CC   decreased such as in osteoporosis.  Disorders characterised by
CC   decreased or increased levels of CTGF-3 can be treated by
CC   administering CTGF-3 polypeptides and anti-CTGF-3 antibodies,
CC   respectively.
SQ   Sequence   1285 BP;    237 A;    418 C;    389 G;    241 T;
```

```
  Query Match             6.2%;  Score 33.8;  DB 1;  Length 1285;
  Best Local Similarity   72.1%;  Pred. No. 0.2;
  Matches   44;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;
```

```
Qy    136 GTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGGAGAAACTTGCTAC 195
          ||||| || | ||||||||||||||| || || ||     ||| ||| ||  | ||| ||
Db    144 GTGCTGGATGGCTGTGGCTGCTGCCGGGTATGTGCACGGCGGCTGGGGGAGCCCTGCGAC 203

Qy    196 C 196
          |
Db    204 C 204
```

RESULT 14
Q36032
ID Q36032 standard; cDNA; 684 BP.
AC Q36032;
DT 24-MAY-1993 (first entry)
DE Chicken nov gene fragment II.
KW avian nephroblastoma; avian myeloblastoma virus;
KW stringent hybridisation; ss.
OS Gallus domesticus.
PN WO9300430-A.
PD 07-JAN-1993.
PF 25-JUN-1992; F00589.
PR 25-JUN-1991; FR-007807.
PA (CNRS) CENT NAT RECH SCI.
PI Martinerie C, Perbal B;
DR WPI; 93-036377/04.
PT Nucleotide sequences hybridising to regions of chicken nov gene -
PT useful as probes for detecting complementary sequences to
PT evaluate development and/or differentiation of tumours
PS Claim 1; Page 26; 67pp; French.
CC The chicken nov gene is stimulated in avian nephroblastoma induced
CC by avian myeloblastoma virus but not in normal adult kidney. A
CC 1975bp cDNA sequence was isolated from a gene bank prepared from
CC chicken embryonic fibroblasts screened with a tumour-derived probe.
CC Fragment II is derived from the nov gene; nucleotide sequences
CC which hybridise to Fragment II under stringent conditions (i.e.
CC 50% formamide, 5 x SCC) are claimed.
SQ Sequence 684 BP; 141 A; 197 C; 226 G; 120 T;

Query Match 6.0%; Score 32.8; DB 1; Length 684;
Best Local Similarity 61.9%; Pred. No. 0.3;
Matches 52; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy    113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
          || || | |   |   |    | ||||| |||| |||  |||||||||  ||||||||
Db     32 GCTGCGCCCCGGGAGTGCCCGCCGTGCTGGACGGCTGCCGCTGCTGCCTGGTGTGCGCCC 91

Qy    173 CAGGGCGGGGAGAAACTTGCTACC 196
             ||| || || |  |||| ||
Db     92 GGCAGCGCGGCGAGAGCTGCTCCC 115
```

RESULT 15
T40206/c
ID T40206 standard; cDNA to mRNA; 1614 BP.
AC T40206;
DT 17-DEC-1996 (first entry)
DE Tobacco lycopene cyclase cDNA.
KW Lycopene cyclase; carotenoid; beta-carotene; herbicide resistance;
KW transgenic plant; crop protection; pigmentation; crtL gene;
KW tobacco; ss.
OS Nicotiana tabacum cv. xanthii.
PN WO9628014-A1.
PD 19-SEP-1996.
PF 05-MAR-1996; U03044.
PR 07-MAR-1995; US-399561.
PA (YISS) YISSUM RES & DEV CO.
PI Cunningham FX, Gantt E, Hirschberg J;
DR WPI; 96-433414/43.
PT DNA encoding lycopene cyclase or mutants resistant to herbicides -
PT used to control carotenoid biosynthesis and to produce herbicide
PT resistant plants and algae
PS Claim 2; Page 39-40; 59pp; English.
CC A cDNA clone (T40206), corresponding to the crtL gene, codes for
CC the lycopene syntase of tobacco cv. xanthii. It was isolated
CC from a leaf cDNA library using the Synechococcus lycopene synthase
CC lcy gene (T40205) as probe. A tomato crtL cDNA (T40207) was
CC similarly isolated. Lycopene synthase genes can be used to control
CC carotenoid biosynthesis in bacteria, algae and plants, e.g. to
CC improve nutritional value (e.g. beta-carotene prodn.),
CC pharmacological properties or appearance. Antisense sequences
CC inhibit synthesis of the enzyme, resulting in increased red
CC pigmentation. Mutant crtL genes can be used to impart herbicide
CC resistance to transgenic plants.
SQ Sequence 1614 BP; 447 A; 288 C; 370 G; 509 T;

Query Match 5.8%; Score 31.8; DB 1; Length 1614;
Best Local Similarity 57.6%; Pred. No. 1;
Matches 57; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy    379 AAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAGTCTTCCAACAGATTTGTT 438
          |||  |  ||    ||| ||||| |||||||  |  ||| | | |       ||| ||| |
Db    816 AAAGGGATGTTCCTCCACTTCTGCCAATATGCCATAAGCTACTTGATATCCAGGTTTATA 757

Qy    439 TCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAAT 477
          |  || |     |   ||| | ||||| |||| | || |
Db    756 TGGCTTATCATACTGAACAAGACATCTTGAGAAGCCAGT 718
```

Search completed: May 3, 1999, 23:14:13
Job time: 1325 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 23:57:01 ; Search time 188.7 Seconds
(without alignments)
4438.039 Million cell updates/sec

Title: US-09-037-460-1_COPY_61_609
Perfect score: 549
Sequence: 1 AAGAGCGTCTTGCTGCTGAC.....GGAAATGGTTAAATCCACGC 549

Scoring table: IDENTITY_NUC

Searched: 2002476 seqs, 762712212 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: gb_est1:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: gb_est6:*
16: gb_est7:*
17: gb_est8:*
18: gb_est9:*
19: gb_est10:*
20: gb_est11:*
21: gb_est12:*
22: gb_est13:*
23: gb_est22:*
24: gb_est14:*
25: gb_est15:*
26: gb_est16:*
27: gb_est17:*
28: gb_est18:*
29: gb_est19:*
30: gb_est20:*
31: gb_est21:*
32: em_est10:*
33: em_est11:*
34: em_est12:*
35: em_est13:*
36: em_est14:*
37: em_est15:*
38: em_est16:*
39: em_est17:*
40: em_est18:*
41: em_est19:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	35.8	6.5	599	23	AA540522	AA540522 LD20232.5
c 2	33.8	6.2	398	10	R48383	R48383 yj67h05.r1
c 3	32.6	5.9	253	31	AU018204	AU018204 Mus muscu
4	32.4	5.9	360	20	AA449580	AA449580 zx06e08.r
c 5	32.2	5.9	242	13	W01086	W01086 za59h04.r1
6	32.2	5.9	355	16	AA168292	AA168292 ms22f08.r
c 7	32	5.8	536	23	AA264358	AA264358 LD08029.5
c 8	32	5.8	546	23	AA439173	AA439173 LD13657.5
c 9	32	5.8	756	23	AA978722	AA978722 LD32729.5
10	31.8	5.8	424	28	AI011272	AI011272 EST205723
11	31.8	5.8	442	11	H06070	H06070 yl72g12.s1
12	31.8	5.8	485	22	AA607213	AA607213 vm84g08.r
c 13	31.6	5.8	394	21	AA485611	AA485611 zx90g08.r
c 14	31.4	5.7	338	27	AA875541	AA875541 UI-R-E0-c
15	31.4	5.7	371	30	AI017046	AI017046 ov01c01.x
16	31.4	5.7	422	10	T80915	T80915 yd25g02.r1
c 17	31.4	5.7	534	23	AA202628	AA202628 LD02785.5
c 18	31	5.6	425	30	AI159468	AI159468 vz75a11.r
19	30.6	5.6	238	30	AI028047	AI028047 ow43a11.x
20	30.6	5.6	296	15	AA027418	AA027418 mi03f07.r
21	30.6	5.6	304	31	AI266608	AI266608 qn29d01.x
22	30.6	5.6	367	28	AA887391	AA887391 oj37a05.s
23	30.6	5.6	400	31	AI262853	AI262853 qk36d04.x
24	30.6	5.6	400	31	AI272063	AI272063 qj89d05.x
25	30.6	5.6	406	27	AA844402	AA844402 aj19e03.s
26	30.6	5.6	443	31	AI264140	AI264140 qk03b09.x
27	30.6	5.6	465	31	AI262554	AI262554 qk42e02.x
28	30.6	5.6	482	30	AI022804	AI022804 ow55e07.x
29	30.6	5.6	489	29	AI088087	AI088087 oz94h04.x
30	30.6	5.6	498	22	AA584389	AA584389 nn77c08.s
c 31	30.4	5.5	250	20	HUML13684	D31399 Human fetal
32	30.6	5.6	564	31	AI193129	AI193129 qe40e11.s
c 33	30.4	5.5	282	17	AA324516	AA324516 EST27355
34	30.4	5.5	502	19	AA434839	AA434839 ve17b08.r
c 35	30.4	5.5	503	23	AI273346	AI273346 ql45g07.x
36	30.4	5.5	546	27	AA196958	AA196958 zq09f07.r
c 37	30.4	5.5	655	23	AA802895	AA802895 GM06533.5
38	30.2	5.5	423	21	AA504577	AA504577 aa60f06.r
39	30.2	5.5	549	14	W85569	W85569 mf58a01.r1
c 40	30.2	5.5	566	29	C79130	C79130 Mus musculu
c 41	30.2	5.5	582	29	C79086	C79086 Mus musculu
c 42	30	5.5	330	17	AA301623	AA301623 EST14659
43	30	5.5	456	14	W79063	W79063 zd80a01.s1
c 44	30	5.5	465	24	AA594191	AA594191 nn29a08.s
45	29.8	5.4	294	25	T20823	T20823 2831 Lambda

ALIGNMENTS

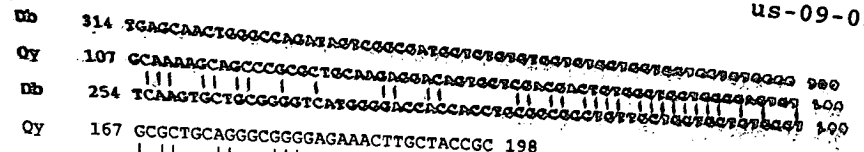
RESULT 1
AA540522/c
LOCUS AA540522 599 bp mRNA EST 28-NOV-1998
DEFINITION LD20232.5prime LD Drosophila melanogaster embryo BlueScript Drosophila melanogaster cDNA clone LD20232 5prime similar to U72331: Drosophila melanogaster putative RNA-binding protein (how) mRNA, complete cds, mRNA sequence.
ACCESSION AA540522
NID g2286956
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 599)
AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE BDGP/HHMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA

```
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 202  row: C  column: 8
            High quality sequence stop: 564.
FEATURES              Location/Qualifiers
     source           1. .599
                      /organism="Drosophila melanogaster"
                      /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                      EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                      Synthesis kit. Oligo dT-primed and directionally cloned at
                      EcoRI and XhoI in BlueScript SK(+/-)"
                      /db_xref="BDGP_EST:BDcln019291"
                      /db_xref="taxon:7227"
                      /clone="LD20232"
                      /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                      /sex="male and female"
                      /dev_stage="0 to 24 hours mixed stage embryonic"
                      /lab_host="SOLR"
BASE COUNT      205 a    164 c    138 g     92 t
ORIGIN


  Query Match             6.5%;  Score 35.8;  DB 23;  Length 599;
  Best Local Similarity   54.1%;  Pred. No. 0.62;
  Matches   73;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      47 TGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGT 106
           ||  |  ||||  ||   | |  ||| ||  |||    |    |||   | || ||| |
Db     536 TGAGCAACTGGGCCAGATAGTCGGCGATGCTCTGTGTGCTCTGCTGCTGCTGCTGTGGGG 477

Qy     107 GCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGT 166
           |||  || ||  |   ||  | ||  |||     || || |||| |||||||| | |  |
Db     476 TCAAGTGCTGCGGGTCATGGAGGACCACACACTGCGGCGCCTGTTGCTGCTGCTGTGCCT 417

Qy     167 GCGCTGCAGGGCGGG 181
           | ||   ||   |||
Db     416 GAGCCTGAGCCTGGG 402


RESULT   2
R48383/c
LOCUS       R48383         398 bp    mRNA            EST       18-MAY-1995
DEFINITION  yj67h05.r1 Homo sapiens cDNA clone 153849 5' similar to gb:M75126
            HEXOKINASE, TYPE I (HUMAN);.
ACCESSION   R48383
NID         g810409
KEYWORDS    EST.
SOURCE      human clone=153849 library=Soares breast 2NbHBst vector=pT7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=M13RP1 Rsite1=Not I Rsite2=Eco RI Adult female.
            1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT 3'],
            double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
            digested with Not I and cloned into the Not I and Eco RI sites of a
            modified pT7T3 vector (Pharmacia). Library went through one round
            of normalization to a Cot = 230. Library constructed by Bento
            Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 257
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .398
                      /organism="Homo sapiens"
                      /clone="153849"
BASE COUNT       70 a    107 c    107 g    111 t      3 others
ORIGIN


  Query Match             6.2%;  Score 33.8;  DB 10;  Length 398;
  Best Local Similarity   48.2%;  Pred. No. 2.4;
  Matches   95;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy     158 GCCGAGTGTGCGCTGCAGGGCGGGGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATG 217
           ||| |  ||| |   |  || |||||| ||   ||||  ||   | | ||  |  |  |
Db     269 GCCCACAGTGGGTGCCCAGGAGGGGAGGAAGCAGCTAATGCTATGCCCCACACTCGCCTT 210

Qy     218 GCATGAAGTGTGGCCCGGGGCTGAGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTG 277
           || | ||||   || || |    |           |||  ||| ||  || | |||
Db     209 TCAGGCAGTGCCCCCGGGAGGGAGGCCCGGCAGTGTCTGCTGGTGATAATACATTTCACA 150

Qy     278 AAGAGTTTGGTATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGA 337
           || |   || | |  ||| ||  | |  |   || |    ||  ||    |  | | ||
Db     149 CAGGGAGGGGGAACCCCAAGGATGGGCTTTGGAGCCCAGAAGGCCTGTCAGGTGGTGTGA 90

Qy     338 CCTGCAACTGCCAGTCA 354
           | | |   |||| ||
Db      89 TTTTCCATCTCCAGCCA 73


RESULT   3
AU018204/c
LOCUS       AU018204       253 bp    mRNA            EST       14-OCT-1998
DEFINITION  Mus musculus 2-cell embryo cDNA 3'-end sequence, clone J0750B07,
            mRNA sequence.
ACCESSION   AU018204
NID         g3373694
KEYWORDS    EST.
SOURCE      Mus musculus (strain:C57BL/6J) 2-cell stage embryo cDNA to mRNA,
            clone:J0750B07.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Ko,M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,
            DePalma,G.E., Liang,Y., Kargul,G.J., Sharara,R. and Doi,H.
  TITLE     Systematic analyses of genes expressed in 2-cell stage mouse
            embryos (The ERATO/Doi Project at Wayne State University)
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 253)
  AUTHORS   Ko,M.S.H. and Doi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1998) to the DDBJ/EMBL/GenBank databases.
            Hirofumi Doi, Japan Science and Technology Corporation (JST), Doi
            Bioasymmetry Project, ERATO; WBG Marive East 12F, 2-6 Nakase,
            Mihama-ku, Chiba, Chiba 261-7112, Japan (E-mail:hd@bioa.jst.go.jp,
            Tel:+81-43-299-1351, Fax:+81-43-297-7530)
FEATURES              Location/Qualifiers
     source           1. .253
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="J0750B07"
                      /dev_stage="2-cell stage embryo"
BASE COUNT       61 a     60 c     65 g     61 t      6 others
```

```
ORIGIN

  Query Match              5.9%;  Score 32.6;  DB 31;  Length 253;
  Best Local Similarity   52.6%;  Pred. No. 5.1;
  Matches   71;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy     359 TCTGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCA 418
           |||  || | |  ||| | || |  |||       |  |||||| |    |  ||| | |
Db     151 TCTTGGAGAAGTTGACAGAAACAGCCCTTTTCCCCGTCTTCTTTCTCCTTGGAGTTAACC 92

Qy     419 AGTCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATA 478
           ||  ||| |     ||| || | ||| |||||  | |  |||    |||     ||||
Db      91 TTTCCCCCACCGTTAATGTGTCCCCCACCGAGCACCAGACTGCAGAAAGAGGCAATAATA 32

Qy     479 TTGTGAGAGAAGAAG 493
           |    | ||||||
Db      31 CGGACCAACAAGAAG 17


RESULT   4
AA449580
LOCUS       AA449580      360 bp    mRNA            EST       04-JUN-1997
DEFINITION  zx06e08.r1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
            785702 5' similar to SW:ACON_BOVIN P20004 ACONITATE HYDRATASE,
            MITOCHONDRIAL PRECURSOR ;, mRNA sequence.
ACCESSION   AA449580
NID         g2163330
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="785702"
                     /clone_lib="Soares total fetus Nb2HF8 9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
     mRNA            <1. .>360
                     /db_xref="GDB:5983320"
BASE COUNT       92 a    106 c    104 g     58 t


ORIGIN

  Query Match              5.9%;  Score 32.4;  DB 20;  Length 360;
  Best Local Similarity   51.4%;  Pred. No. 6.3;
  Matches   75;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     206 CAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTGAGGTGTCAGCCTTCTAATGGGGAGG 265
           | |||| |||||||| |||||   | | || | |  ||   |||          |||
Db      19 CGGGCACGGATGGCAAGAAGTTCAGGCTGGAGGTCCGGATGCAGATGAGCTTCCCAAAGG 78

Qy     266 ATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGG 325
           | ||||    || |   |  | || | |  ||   ||| | | || |  ||||  |
Db      79 GGACGTTTGACCCAGGGCAGGACACCTACCAGCACCCACCCAAGGACAGCAGCGGGCAGC 138

Qy     326 ATTGCAGAGAGACCTGCAACTGCCAG 351
           ||      | || |  || | |||||
Db     139 ATGTGGACGTGAGCCCCACCAGCCAG 164


RESULT   5
W01086/c
LOCUS       W01086        242 bp    mRNA            EST       18-APR-1996
DEFINITION  za59h04.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            296887 5', mRNA sequence.
ACCESSION   W01086
NID         g1273230
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 176.
FEATURES             Location/Qualifiers
     source          1. .242
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="296887"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>242
BASE COUNT       60 a     50 c     57 g     73 t      2 others
ORIGIN
```

```
  Query Match              5.9%;  Score 32.2;  DB 13;  Length 242;
  Best Local Similarity   61.2%;  Pred. No. 6.7;
  Matches   52;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy    422 CTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATTG 481
          ||||   |  | |  | | |||||  || ||| ||  |||| |   | ||  | | ||
Db    187 CTTCGCTCCAAATATTCTTTCTCAATTAGAATGTCAATGCATTTCTGGGTGTTATTTTTA 128

Qy    482 TGAGAGAAGAAGTTGTGAAAGAGAA 506
            |  ||||| || ||||||||||||
Db    127 AAAAGGAAGAGGTAGTGAAAGAGAA 103


RESULT   6
AA168292
LOCUS       AA168292      355 bp    mRNA            EST       12-FEB-1997
DEFINITION  ms22f08.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            607719 5' similar to gb:X03212 KERATIN, TYPE II CYTOSKELETAL 7
            (HUMAN); gb:X03491 Mouse mRNA for type II 57 kd keratin (MOUSE);,
            mRNA sequence.
ACCESSION   AA168292
NID         g1746494
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:373151
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 329.
FEATURES             Location/Qualifiers
     source          1. .355
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. Whole
                     skin from 11 week old C57BL/6 female mice. Average insert
                     size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="607719"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
     mRNA            <1. .>355
BASE COUNT       100 a     76 c     119 g     60 t
ORIGIN


  Query Match              5.9%;  Score 32.2;  DB 16;  Length 355;
  Best Local Similarity   48.6%;  Pred. No. 7.2;
  Matches   88;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy      2 AGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCA 61
          ||| |    |||||| ||||| |        |||     ||| |||        |    |||
Db     10 AGAACCAGGTGCTGCAGACCAAGTGGGAGCTGCTGCAGCACGTGGACCTGAACAACTGCA 69

Qy     62 ATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGC 121
          | ||  |   ||    |  || | |   |     |||||    | | || ||| |    |
Db     70 AGAACAACCTGGAACCCATCCTTGAGGGCCACATCAGCAACATGCGGAACAGCTGGAGAC 129

Qy    122 GCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGG 181
          ||||   | ||||||  |   | |     |    ||||  || ||||||   |  || ||
Db    130 GCTGTCTGGGGACAGAGTGAGGCTGGACTCTGAGCTGAGGAATGTGCGTGACGTGGTGGA 189

Qy    182 G 182
          |
Db    190 G 190


RESULT   7
AA264358/c
LOCUS       AA264358      536 bp    mRNA            EST       29-NOV-1998
DEFINITION  LD08029.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD08029 5prime, mRNA sequence.
ACCESSION   AA264358
NID         g1900427
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Blast expect value = 3.7E-211 on U72331: Drosophila melanogaster
            putative RNA-binding protein (how) mRNA, complete cds
            Plate: 80  row: C  column: 5
            High quality sequence stop: 535.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Drosophila melanogaster"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
                     /db_xref="BDGP_EST:BDcln007384"
                     /db_xref="taxon:7227"
                     /clone="LD08029"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
BASE COUNT       143 a     171 c     138 g      84 t
ORIGIN


  Query Match              5.8%;  Score 32;  DB 23;  Length 536;
  Best Local Similarity   50.7%;  Pred. No. 9;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     47 TGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGT 106
          ||  |  ||||  ||   | |  ||| ||  |||    |    |||   | || ||| |
```

BEST AVAILABLE COPY



```
Qy     167 GCGCTGCAGGGCGGGGAGAAACTTGCTACCGC 198
           | ||   ||   |||      ||||   | ||
Db     194 GAGCCTGAGCCTGGGCTTGGGCTTGAGCCTGC 163
```

```
RESULT   8
AA439173/c
LOCUS       AA439173      546 bp    mRNA            EST       27-NOV-1998
DEFINITION  LD13657.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD13657 5prime similar to
            U72331: Drosophila melanogaster putative RNA-binding protein (how)
            mRNA, complete cds, mRNA sequence.
ACCESSION   AA439173
NID         g2793228
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 20, 1998 this sequence version replaced gi:2151054.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 136  row: E  column: 9
            High quality sequence stop: 431.
FEATURES                Location/Qualifiers
     source             1. .546
                        /organism="Drosophila melanogaster"
                        /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                        EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                        Synthesis kit. Oligo dT-primed and directionally cloned at
                        EcoRI and XhoI in BlueScript SK(+/-)"
                        /db_xref="BDGP_EST:BDcln012884"
                        /db_xref="taxon:7227"
                        /clone="LD13657"
                        /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                        /sex="male and female"
                        /dev_stage="0 to 24 hours mixed stage embryonic"
                        /lab_host="SOLR"
BASE COUNT      196 a    147 c    126 g     77 t
ORIGIN
```

```
  Query Match              5.8%;  Score 32;  DB 23;  Length 546;
  Best Local Similarity   50.7%;  Pred. No. 9.1;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy      47 TGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGT 106
           ||  |  ||||  ||   | |  ||| ||  |||     |    |||   | || ||| |
Db     526 TGAGCAACTGGGCCAGATAGTCGGCGATGCTCTGTGTGCTCTGCTGCTGCTGCTGTGGGG 467

Qy     107 GCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGT 166
           |||  || ||  |  |      ||  ||       || || |||| |||||||| | |  |
Db     466 TCAAGTGCTGCGGGGTCATGGGGACCACCACCTGCGGCGCCTGTTGCTGCTGCTGTGCCT 407

Qy     167 GCGCTGCAGGGCGGGGAGAAACTTGCTACCGC 198
           | ||   ||   |||      ||||   | ||
```



```
LOCUS       AA978722      780 bp    mRNA            EST       24-NOV-1998
DEFINITION  LD32729.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD32729 5prime similar to U85651;
            Drosophila melanogaster KH RNA binding protein (who) mRNA, complete
            cds, mRNA sequence.
ACCESSION   AA978722
NID         g3156138
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 756)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 327  row: C  column: 5
            High quality sequence stop: 753.
FEATURES                Location/Qualifiers
     source             1. .756
                        /organism="Drosophila melanogaster"
                        /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                        XhoI; Sized fractionated cDNAs were directly ligated into
                        pOT2. "
                        /db_xref="taxon:7227"
                        /clone="LD32729"
                        /clone_lib="LD Drosophila melanogaster embryo pOT2"
                        /sex="male and female"
                        /dev_stage="0 to 24 hours mixed stage embryonic"
                        /lab_host="XL1 Blue"
BASE COUNT      238 a    216 c    182 g    120 t
ORIGIN
```

```
  Query Match              5.8%;  Score 32;  DB 23;  Length 756;
  Best Local Similarity   50.7%;  Pred. No. 9.6;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy      47 TGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGT 106
           ||  |  ||||  ||   | |  ||| ||  |||     |    |||   | || ||| |
Db     504 TGAGCAACTGGGCCAGATAGTCGGCGATGCTCTGTGTGCTCTGCTGCTGCTGCTGTGGGG 445

Qy     107 GCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGT 166
           |||  || ||  |  |      ||  ||       || || |||| |||||||| | |  |
Db     444 TCAAGTGCTGCGGGGTCATGGGGACCACCACCTGCGGCGCCTGTTGCTGCTGCTGTGCCT 385

Qy     167 GCGCTGCAGGGCGGGGAGAAACTTGCTACCGC 198
           | ||   ||   |||      ||||   | ||
Db     384 GAGCCTGAGCCTGGGCTTGGGCTTGAGCCTGC 353
```

```
RESULT  10
AI011272
LOCUS       AI011272      424 bp    mRNA            EST       15-JUN-1998
DEFINITION  EST205723 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVAS70 3' end, mRNA sequence.
ACCESSION   AI011272
NID         g3225104
KEYWORDS    EST.
```

```
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source           1. .424
                      /organism="Rattus sp."
                      /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
                      /db_xref="taxon:10118"
                      /clone="ROVAS70"
                      /clone_lib="Normalized rat ovary, Bento Soares"
BASE COUNT       86 a    112 c    123 g    103 t
ORIGIN
```

Query Match 5.8%; Score 31.8; DB 28; Length 424;
Best Local Similarity 57.6%; Pred. No. 10;
Matches 57; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy    195 CCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTGAGGTGTCAGCCTTC 254
          ||  || | | |||| |    |||||     |  |||   ||||||||| |||| |   |
Db    230 CCTGACGGGCCCAGGAAGACTTGGCAAATGTTAGGGCTGTGGGCTGAGGGGTCACCAGGC 289

Qy    255 TAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTG 293
            || |  | ||| | ||  ||  |||  | ||| |  ||
Db    290 CAAAGCTGGGGAACTTTGGGGCAAAGGCTGTGGCACATG 328
```

```
RESULT  11
H06070
LOCUS       H06070        442 bp    mRNA            EST       21-JUN-1995
DEFINITION  yl72g12.s1 Homo sapiens cDNA clone 43492 3'.
ACCESSION   H06070
NID         g869622
KEYWORDS    EST.
SOURCE      human clone=43492 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=Promega -21m13 Rsite1=Not
            I Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-416-033
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 354
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .442
                      /organism="Homo sapiens"
                      /clone="43492"
BASE COUNT      113 a     78 c     91 g    148 t     12 others
ORIGIN
```

Query Match 5.8%; Score 31.8; DB 11; Length 442;
Best Local Similarity 53.7%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```
Qy    201 AGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTGAGGTGTCAGCCTTCTAATGG 260
          || || ||| ||   || || |    |  | |  | | || |   ||||    | | |||
Db    240 AGACTGAGGTATCCGTGACAGGGCAGGAAGGCATGTGATGGGTCCTCAGTTCCCAAGTGG 299

Qy    261 GGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGACTGTCCCTACGGCACCTTCGG 320
          | ||||  ||  |||   | ||||     || | || | || || | |   |  ||| |
Db    300 GAAGGAGGCTAATGGGTTAAAGTTCATATTCAGGAAGGGCTTTCACATCCATAGTTTCTG 359

Qy    321 GAT 323
          |||
Db    360 GAT 362
```

```
RESULT  12
AA607213
LOCUS       AA607213      485 bp    mRNA            EST       30-SEP-1997
DEFINITION  vm84g08.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone 1004990 5', mRNA sequence.
ACCESSION   AA607213
NID         g2456106
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:569206
            High quality sequence stop: 466.
FEATURES              Location/Qualifiers
     source           1. .485
```

```
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 800 blastocysts.   Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.  cDNAs  were
                     cloned into the NotI/SalI sites of a pSPORT vector (Life
                     Technologies).  Two different size selections:  B1 (larger
                      inserts) and B3."
                     /db_xref="taxon:10090"
                     /clone="1004990"
                     /clone_lib="Knowles Solter mouse blastocyst B1"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"
BASE COUNT       144 a    102 c    124 g    115 t
ORIGIN
```

```
  Query Match             5.8%;  Score 31.8;  DB 22;  Length 485;
  Best Local Similarity   52.7%;  Pred. No. 10;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      16 CTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAATAATTATGCGGTG 75
           |||| || | |   |    ||| |  ||  ||||  |   ||    |  |  |      |
Db     216 CTGAACATGGTGTACACATCTGAAATCCCAGTGGGTGGGAGGTAAAGGCAAGAGAATCAG 275

Qy      76 GACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCGCTGCAAGAGGACA 135
           || |||  |  |  ||  |      ||||||||||   |||||| ||||||||||   ||
Db     276 GAATGCAGTTTAGCCTTGGCTGCATGTGAGTGCAATGCCAGCCCGGGCTGCAAGAAACCA 335

Qy     136 GTGCTCGACGA 146
           |  |   | ||
Db     336 GATCCTAATGA 346
```

```
RESULT  13
AA485611/c
LOCUS       AA485611      394 bp    mRNA            EST       11-AUG-1997
DEFINITION  zx90g08.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 811070
            5', mRNA sequence.
ACCESSION   AA485611
NID         g2214830
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 385.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Homo sapiens"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="811070"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>394
                     /db_xref="GDB:6041982"
BASE COUNT       79 a    126 c    125 g     64 t
ORIGIN
```

```
  Query Match             5.8%;  Score 31.6;  DB 21;  Length 394;
  Best Local Similarity   58.5%;  Pred. No. 11;
  Matches   55;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     177 GCGGGGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGG 236
           |||| | | |     |    | |  | ||||   || ||| |||| || ||| ||||||||
Db      94 GCGGTGGGCACTGGGACCGCTCCAATTCTCCCACAGGGACGGCAGGATGTGAGGCCCGGG 35

Qy     237 GCTGAGGTGTCAGCCTTCTAATGGGGAGGATCCT 270
           ||||  |      ||  |  | ||||  |||| |
Db      34 GCTGCAGCACTCACCACCCCACGGGGTTGATCTT 1
```

```
RESULT  14
AA875541/c
LOCUS       AA875541      338 bp    mRNA            EST       20-MAR-1998
DEFINITION  UI-R-E0-cv-g-12-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-cv-g-12-0-UI 3' similar to gi|2863898|gb|AA800943|AA800943
            EST190440 Normalized rat lung, Bento Soares Rattus sp. cDNA 3' end,
            mRNA sequence.
ACCESSION   AA875541
NID         g2980489
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from 8, 12 and 18-day embryo.  The
                     tag is a string  of 3-5 nucleotides present between the
```

Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
/db_xref="taxon:10116"
/clone="UI-R-E0-cv-g-12-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
BASE COUNT 61 a 112 c 132 g 33 t
ORIGIN

Query Match 5.7%; Score 31.4; DB 27; Length 338;
Best Local Similarity 52.7%; Pred. No. 13;
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy      70 GCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCGCTGCAAG 129
           |||| |   || |||   | ||| |||         ||    |||   |   ||     |
Db     129 GCGGCGCCTTGTCCTGCGCGCTGCGACGTGTCGCGCTGTCCGAGCCCTCGCTGCCCTGGG 70

Qy     130 AGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGGAGAAACT 189
            |    ||||  |||  |||   |||||||  |||||||| ||  |   ||| ||  |
Db      69 GGCTATGTGCCAGACCTCTGCAACTGCTGCTTGGTGTGCGCCGCCAGCGAGGGTGAGCCC 10

Qy     190 TGCTACCGC 198
           |||  ||||
Db       9 TGCGGCCGC 1
```

RESULT 15
AI017046
LOCUS AI017046 371 bp mRNA EST 27-AUG-1998
DEFINITION ov01c01.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1636032 3', mRNA sequence.
ACCESSION AI017046
NID g3231382
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 371)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1051 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 350.
FEATURES Location/Qualifiers
source 1. .371
/organism="Homo sapiens"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1636032"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
BASE COUNT 76 a 99 c 119 g 77 t
ORIGIN

Query Match 5.7%; Score 31.4; DB 30; Length 371;
Best Local Similarity 49.1%; Pred. No. 13;
Matches 83; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
Qy     162 AGTGTGCGCTGCAGGGCGGGGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCAT 221
           | |  | ||||   || ||||       |||     | ||| ||| | | |   |    |
Db     139 ACTCAGTGCTGAGTGGAGGGGGATGGGGTGCCGGATCTCAGGCTCTGCCTTCCTTTTGTT 198

Qy     222 GAAGTGTGGCCCGGGGCTGAGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGA 281
             ||  ||| | |   |||| | ||||   ||   || |  ||| ||   |  | |||
Db     199 CGGATGGAGCCTGAGAGAGAGGGGACAGCTGGCTCCAGGCGGAGATGCTGAAGCAGCAGA 258

Qy     282 GTTTGGTATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGGATTGC 330
           |  || |  || ||  |||    ||  |    |  | |||| |  | ||
Db     259 GGGTGCTGGCTCCACTGACCACTCCACCCCGTCTGTGGGGAGGCGTGGC 307
```

Search completed: May 4, 1999, 00:11:42
Job time: 881 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 14:59:21 ; Search time 398.05 Seconds
(without alignments)
4934.239 Million cell updates/sec

Title: US-09-037-460-1_COPY_61_609
Perfect score: 549
Sequence: 1 AAGAGCGTCTTGCTGCTGAC.....GGAAATGGTTAAATCCACGC 549

Scoring table: IDENTITY_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl1:*
9: gb_pl2:*
10: gb_pr1:*
11: gb_pr2:*
12: gb_pr3:*
13: gb_ro:*
14: gb_st:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: gb_htg:*
19: em_ba:*
20: em_fun:*
21: em_hum1:*
22: em_hum2:*
23: em_in:*
24: em_om:*
25: em_or:*
26: em_ov:*
27: em_pat:*
28: em_ph:*
29: em_pl:*
30: em_ro:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: em_htg:*
35: em_sts:*
36: gb_ba1:*
37: gb_ba2:*
38: gb_pl1:*
39: gb_pl2:*
40: gb_pr1:*
41: gb_pr2:*
42: gb_pr3:*
43: gb_sts:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	549	100.0	1271	6	AR004664	AR004664 Sequence
2	549	100.0	2006	10	HSRNAESM1	X89426 H.sapiens m
3	549	100.0	2006	40	HSRNAESM1	X89426 H.sapiens m
4	340.8	62.1	2009	13	RNU80818	U80818 Rattus norv
5	38	6.9	6504	3	AAU02548	U02548 Aedes aegyp
6	38	6.9	8780	3	MQSVITE	L41842 Aedes aegyp
c 7	38.2	7.0	38962	1	SC2E1	AL023797 Streptomy
c 8	38.2	7.0	38962	36	SC2E1	AL023797 Streptomy
c 9	37.4	6.8	162485	18	AC004840	AC004840 *** SEQUE
10	36.4	6.6	2267	13	MUSFISP12B	M70642 Mouse FISP-
11	36.4	6.6	2330	13	MUSTGFB	M80263 Mouse mRNA
12	36.4	6.6	4128	13	MUSFISP12A	M70641 Mouse FISP-
13	36	6.6	1927	5	CJU13063	U13063 Coturnix ja
14	35.8	6.5	1975	6	A28405	A28405 Sequence I
15	35.8	6.5	1975	6	A28444	A28444 nov mRNA se
16	35.8	6.5	1976	5	GGNOVMRNA	X59284 G.gallus no
17	36	6.6	40397	8	SC9959	Z49939 S.cerevisia
18	36	6.6	40397	38	SC9959	Z49939 S.cerevisia
19	35.4	6.4	1406	8	CRU24500	U24500 Chlamydomon
20	35.4	6.4	1406	38	CRU24500	U24500 Chlamydomon
c 21	35.4	6.4	2154	5	TMCRETRA	X77306 T.marmorata
22	35.2	6.4	2075	6	AR018957	AR018957 Sequence
23	35.2	6.4	2075	6	I11636	I11636 Sequence 1
24	35.2	6.4	2075	6	I32210	I32210 Sequence 1
25	35.2	6.4	2075	10	HUMCONGRO	M92934 Human conne
26	35.2	6.4	2075	40	HUMCONGRO	M92934 Human conne
27	35.2	6.4	2312	11	HSCTGF	X78947 H.sapiens m
28	35.2	6.4	2312	41	HSCTGF	X78947 H.sapiens m
29	35	6.4	1496	4	SSU83916	U83916 Sus scrofa
30	34.4	6.3	230	6	A28407	A28407 Sequence th
31	34.4	6.3	684	6	A28406	A28406 Sequence II
c 32	35	6.4	63033	1	MTV008	AL021246 Mycobacte
c 33	35	6.4	63033	36	MTV008	AL021246 Mycobacte
34	34.4	6.3	10471	3	AF008228	AF008228 Drosophil
35	34.4	6.3	11451	3	DMTENM	X73154 D.melanogas
36	34.4	6.3	14045	3	DMODZ4	AF008227 Drosophil
37	33.8	6.2	1283	12	AF083500	AF083500 Homo sapi
38	33.8	6.2	1283	42	AF083500	AF083500 Homo sapi
39	33.8	6.2	1309	11	AF074604	AF074604 Homo sapi
40	33.8	6.2	1309	41	AF074604	AF074604 Homo sapi
c 41	33	6.0	4180	13	RNU49058	U49058 Rattus norv
42	32.8	6.0	1781	13	MUSH1X	L26164 Mouse histo
43	33.6	6.1	323900	18	HS402G11	AL022328 Human DNA
c 44	32.6	5.9	1028	5	XLGAP	Y00791 Xenopus lae
45	32.8	6.0	5597	13	MMHIST431	Y12290 M.musculus

ALIGNMENTS

RESULT 1
AR004664
LOCUS AR004664 1271 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5747280.
ACCESSION AR004664
NID g3965543
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1271)
AUTHORS Hastings,G.A. and Rosen,C.A.
TITLE Human vascular IBP-like growth factor
JOURNAL Patent: US 5747280-A 1 05-MAY-1998;
FEATURES Location/Qualifiers
source 1. .1271
/organism="unknown"
BASE COUNT 361 a 246 c 336 g 320 t 8 others
ORIGIN

Query Match 100.0%; Score 549; DB 6; Length 1271;
Best Local Similarity 100.0%; Pred. No. 3.2e-156;

```
  Matches  549;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120

Qy     61 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180

Qy    121 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240

Qy    181 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300

Qy    241 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360

Qy    301 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420

Qy    361 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480

Qy    421 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540

Qy    481 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600

Qy    541 AATCCACGC 549
          |||||||||
Db    601 AATCCACGC 609


RESULT   2
HSRNAESM1
LOCUS       HSRNAESM1    2006 bp    RNA             PRI       07-OCT-1996
DEFINITION  H.sapiens mRNA for ESM-1 protein.
ACCESSION   X89426
NID         g1150418
KEYWORDS    ESM-1 protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2006)
  AUTHORS   Lassalle,P., Molet,S., Janin,A., Heyden,J.V., Tavernier,J.,
            Fiers,W., Devos,R. and Tonnel,A.B.
  TITLE     ESM-1 is a novel human endothelial cell-specific molecule expressed
            in lung and regulated by cytokines
  JOURNAL   J. Biol. Chem. 271 (34), 20458-20464 (1996)
  MEDLINE   96355375
REFERENCE   2  (bases 1 to 2006)
  AUTHORS   Lassalle,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1995) P.M. Lassalle, INSERM, Unite 416, 1, bd du
            Prof. CALMETTE, LILLE 59019, FRANCE
FEATURES             Location/Qualifiers
     source          1. .2006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HUVEC"
                     /clone="A11.1"
     sig_peptide     56. .112
     CDS             56. .610
                     /codon_start=1
                     /product="ESM-1 secretory protein"
                     /db_xref="PID:e189266"
                     /db_xref="PID:g1150419"
                     /translation="MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK
                     RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDC
                     PYGTFGMDCRETCNCQSGICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGN
                     IVREEVVKENAAGSPVMRKWLNPR"
BASE COUNT      623 a    333 c    475 g    575 t
ORIGIN


  Query Match           100.0%;  Score 549;  DB 10;  Length 2006;
  Best Local Similarity  100.0%;  Pred. No. 3.4e-156;
  Matches  549;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     59 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 118

Qy     61 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    119 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 178

Qy    121 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 238

Qy    181 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG.240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    239 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 298

Qy    241 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 358

Qy    301 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 418

Qy    361 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 478

Qy    421 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 538

Qy    481 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 598

Qy    541 AATCCACGC 549
          |||||||||
Db    599 AATCCACGC 607


RESULT   3
HSRNAESM1
LOCUS       HSRNAESM1    2006 bp    RNA             PRI       07-OCT-1996
DEFINITION  H.sapiens mRNA for ESM-1 protein.
ACCESSION   X89426
NID         g1150418
KEYWORDS    ESM-1 protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2006)
  AUTHORS   Lassalle,P., Molet,S., Janin,A., Heyden,J.V., Tavernier,J.,
            Fiers,W., Devos,R. and Tonnel,A.B.
  TITLE     ESM-1 is a novel human endothelial cell-specific molecule expressed
```

```
          in lung and regulated by cytokines
  JOURNAL   J. Biol. Chem. 271 (34), 20458-20464 (1996)
  MEDLINE   96355375
REFERENCE   2  (bases 1 to 2006)
  AUTHORS   Lassalle,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1995) P.M. Lassalle, INSERM, Unite 416, 1, bd du
            Prof. CALMETTE, LILLE 59019, FRANCE
FEATURES             Location/Qualifiers
     source          1. .2006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HUVEC"
                     /clone="A11.1"
     sig_peptide     56. .112
     CDS             56. .610
                     /codon_start=1
                     /product="ESM-1 secretory protein"
                     /db_xref="PID:e189266"
                     /db_xref="PID:g1150419"
                     /translation="MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK
                     RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDC
                     PYGTFGMDCRETCNCQSGICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGN
                     IVREEVVKENAAGSPVMRKWLNPR"
BASE COUNT      623 a    333 c    475 g    575 t
ORIGIN
```

Query Match 100.0%; Score 549; DB 40; Length 2006;
Best Local Similarity 100.0%; Pred. No. 3.4e-156;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 60
Db      59 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 118

Qy      61 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 120
Db     119 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 178

Qy     121 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 180
Db     179 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 238

Qy     181 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 240
Db     239 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 298

Qy     241 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 300
Db     299 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 358

Qy     301 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 360
Db     359 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 418

Qy     361 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 420
Db     419 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 478

Qy     421 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 480
Db     479 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 538

Qy     481 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 540
Db     539 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 598

Qy     541 AATCCACGC 549
Db     599 AATCCACGC 607
```

```
RESULT   4
RNU80818
LOCUS       RNU80818     2009 bp    mRNA            ROD       23-DEC-1996
DEFINITION  Rattus norvegicus pineal specific PG25 mRNA, complete cds.
ACCESSION   U80818
NID         g1750197
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2009)
  AUTHORS   Wang,X., Brownstein,M.J. and Young,W.S.
  TITLE     PG25, a full-length pineal-specific cDNA, cloned by differential
            display PCR (DDPCR) and rapid amplification of cDNA ends (RACE)
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 2009)
  AUTHORS   Wang,X., Brownstein,M.J. and Young,W.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1996) LCMR, NIMH, 36 Convent Drive, MSC 40688,
            Bethesda, MD 20892-4068, USA
FEATURES             Location/Qualifiers
     source          1. .2009
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /tissue_type="pineal gland"
     CDS             51. .605
                     /note="similar to human lung endothelial cell-specific
                     ESM-1 gene, encoded by Genbank Accession Number X89426"
                     /codon_start=1
                     /product="PG25"
                     /db_xref="PID:g1750198"
                     /translation="MKSLLLLTTLLIPLHLGMAWSAKYAVDCPEHCDNTECRSSLRCK
                     RTVLDDCGCCQVCAAGPGETCYRTVSGMDGVKCGPGLKCHFYSEEDDFGDEFGVCKDC
                     PYGTFGMDCKETCNCQSGICDRVTGRCLDFPFFQYAAAKSPSRTSASQTERDAASGDG
                     NAVREEIGDRNAARPSVMKWLNPR"
BASE COUNT      576 a    346 c    503 g    584 t
ORIGIN
```

Query Match 62.1%; Score 340.8; DB 13; Length 2009;
Best Local Similarity 78.6%; Pred. No. 5.2e-93;
Matches 434; Conservative 0; Mismatches 112; Indels 6; Gaps 2;

```
Qy       1 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 60
Db      54 AAGAGCCTCTTGCTACTGACCACGCTCCTGATTCCTCTGCACCTGGGCATGGCCTGGAGC 113

Qy      61 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 120
Db     114 GCCAAGTATGCGGTGGATTGCCCGGAACATTGTGACAATACTGAGTGCAGAAGCAGCCTG 173

Qy     121 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 180
Db     174 CGTTGCAAGAGGACAGTGCTGGATGACTGTGGCTGCTGCCAAGTGTGCGCTGCGGGGCCT 233

Qy     181 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 240
Db     234 GGGGAAACCTGCTACCGTACAGTCTCAGGCATGGACGGTGTCAAGTGTGGTCCGGGACTG 293

Qy     241 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 300
Db     294 AAGTGTCACTTTTACAGCGAGGAAGATGATTTCGGTGACGAGTTTGGTGTCTGTAAAGAC 353

Qy     301 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 360
Db     354 TGTCCCTACGGCACATTCGGGATGGATTGCAAGGAGACTTGCAATTGCCAGTCGGGCATA 413

Qy     361 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 420
Db     414 TGTGACAGGGTGACCGGGAGGTGTCTGGACTTCCCCTTCTTCCAGTATGCAGCAGCCAAG 473
```

```
Qy     421 TCTTCCAACAG---ATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAAT 477
           ||  ||| |||    | ||  || |  || |||| ||    || ||||||||||||||||
Db     474 TCGCCCAGCAGGACCTCTGCCTCCCAAACAGAGCGCGATGCAGCCTCTGGAGATGGCAAT 533

Qy     478 ATTGTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGG 537
           ||||||||||||  |||   |  ||| |||||| || |  | |||||   |||||||
Db     534 GCTGTGAGAGAAGAGATTGGAGACCGGAACGCTGCCCGGCCCTCTGTAAT---GAAATGG 590

Qy     538 TTAAATCCACGC 549
           |||||||| |||
Db     591 TTAAATCCGCGC 602


RESULT   5
AAU02548
LOCUS       AAU02548     6504 bp    mRNA            INV       20-MAY-1994
DEFINITION  Aedes aegypti UGALS vitellogenin mRNA, complete cds.
ACCESSION   U02548
NID         g472307
KEYWORDS    .
SOURCE      yellow fever mosquito.
  ORGANISM  Aedes aegypti
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Aedes.
REFERENCE   1  (bases 1 to 6504)
  AUTHORS   Chen,J.S., Cho,W.L. and Raikhel,A.S.
  TITLE     Analysis of mosquito vitellogenin cDNA: similarity with vertebrate
            phosvitins and arthropod serum proteins
  JOURNAL   J. Mol. Biol. 237, 641-647 (1994)
  MEDLINE   94210487
REFERENCE   2  (bases 1 to 6504)
  AUTHORS   Chen,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1993) Chen J., Michigan State University,
            Genetics Program, S-138 Plant Biology Bldg., East Lansing, MI
            48824, USA
FEATURES             Location/Qualifiers
     source          1. .6504
                     /organism="Aedes aegypti"
                     /strain="UGALS"
                     /db_xref="taxon:7159"
                     /clone_lib="Lambda Zap-II Vitellogenic fat body cDNA
                     library"
                     /sex="female"
                     /tissue_type="fat bodies"
                     /dev_stage="adult"
     CDS             10. .6429
                     /note="6 wobble bases due to individual variation:
                     #102(G/A), #123 (G/A), #144 (T/C), #225 (T/C), #231 (C/T),
                     #240 (C/T)"
                     /codon_start=1
                     /product="vitellogenin"
                     /db_xref="PID:g472308"
                     /translation="MLAKLLLLALAGLTAAYQYENSFKGYNPGYKGYDAGYKGYGYDA
                     GYKYNNQGYSYKNGFEYGYQNAYQAAFYKHRPNVTEFEFSSWMPNYEYVYNVTSKTMT
                     ALAELDDQWTGVFTRAYLVIRPKSRDYVVAYVKQPEYAVFNERLPHGYATKFYHDMFK
                     FQPMPMSSKPFGIRYHKGAIKGLYVEKTIPNNEVNILKAWISQLQVDTRGANLMHSSK
                     PIHPSKNEWNGHYKVMEPLVTGECETHYDVNLIPAYMIQAHKQWVPQGQLRGEDDQFI
                     QVTKTQNFDRCDQRMGYHFGFTGYSDFRPNTNQMGNVASKSLVSYMYLTGNWYNFTIQ
                     SSSMINKVAIAPSLVNKEPALVYAQVNMTLNDVHPYDKVPMGPAEDLKVFVDLVYSYN
                     MPSDKKNYVRPGNETSSSSSSSSSSSSSSESSSSSSESVENPKISPVEQYKPLLDKV
                     EKRGNRYRRDLNAIKEKKYYEAYKMDQYRLHRLNDTSSDSSSSDSSSSSSSESKEHRN
                     GTSSYSSSSSSSSSSSSSESSSYSSSSSSSESYSISSEEYYYQPTPANFSYAPEAPF
                     LPFFTGYKGYNIFYARNVDAIRSVGKLVEEIASDLENPSNLPKANTMSKFNILTRAIR
                     AMGYEDIYELAQKYFVSQKERQVAQFSDKKFSKRVDAWVTLRDAVAEAGTPSAFKLIF
                     DFIKEKKLRGYEAATVIASLAQSIRYPTEHLLHEFFLLVTSDVVLHQEYLNATALFAY
                     SNFVNQAHVSNRSAYNYYPVFSFGRLADADYKIIEHKIVPWFAHQLREAVNEGDSVKI
                     QVYIRSLGNLGHPQILSVFEPYLEGTIQITDFQRLAIMVALDNLVIYYPSLARSVLYR
                     AYQNTADVHEVRCAAVHLLMRTDPPADMLQRMAEFTHHDPSLYVRAAVKSAIETAALA
                     DDYDEDSKLALNAKAAINFLNPEDVSIQYSFNHIRDYALENLELSYRLHYGEIASNDH
                     RYPSGLFYHLRQNFGGFKKYTSFYYLVSSMEAFFDIFKKQYNTKYFADYYKSADYSTN
                     YYNFDKYSKYYKQYYYSKDSEYYQKFYGQKKDYYNDKEPFKFTAPRIAKLLNIDAEEA
                     EQLEGQLLFKLFNGYFFTAFDNQTIENLPHKMRHLFENLEDGYAFDVTKFYQQQDVVL
                     AWPLATGFPFIYTLKAPTVFKFEVDASAKTHPQVYKMPAGHPETENDDFFYMPQSING
                     SVDVNLLYHRMVDAKVGFVTPFDHQRYIAGYQKKLHGYLPFNVELGLDFVKDEYEFEF
                     KFLEPKDDHLLFHMSSWPYTGYKDITDMRPIAENPNAKIVHDDNQSTKTMEHTFGQDM
                     TGVALRFHAKYDFDLINFQQFWSLIQKNDFVSAVNYPFAYQPYEYHQFNLFYDSQRTH
                     AKSFKFYAYQKFGAPSFEETGPKHPANRHSYSGNYYESNYAQPFVYSPGSQRRYEQFF
                     RNAASGIRNSFVRYYDFGFEFYAPQYKSEFTFTTAFADSPVDKTSRQLYYFYASPMFP
                     SQSYFKDIPFSGKQFQFCATATSEFPRVPYLKFSDFDKYYGDASQYFDFLYGESCQGG
                     AHIAVKGKQKQTGKYREYLRFSDVAKACKEQMANGYYQFEECQQAIDQAYYYDFYDYA
                     IEYKDVGSVAKNLTNKFYNYFQYAFYPYFESNFFYHGKSNYIKAEFEFAPYGDYYNAS
                     FFGPSYAFQVQNYPVFNDYSTYFPYFFKYTFFPRYQPYYMHRLPSHKPRNRPYYELSN
                     YEQFAIFDRKPQYPSCSFSNDYFYTFDNKKYFYDMGECWHAVMYTVKPDYDFYAQQSH
                     FYNSDFEYKYKNGFEEYEQFAALARRGSDNQLYFKFLFGDNYIEVFPNNGGVPFVKYN
                     GRPYDISKSNIAHFEYKEGYPSFPFFYAFAYPNKDLEVSFFGGKLKFATDGYRARFFS
                     DYSFYNNFVGLCGTNNGEYFDEFVTPDQCYMRKPEFFAASYAITGQNCTGPAKAFNYA
                     YQQKAKQECVKREVYYGDIIYNQEYYHPRYRYYNHNVEESSSSSSSSSSDSSSSSSSS
                     EFSSLSRSGSSSSSSSSEEQKEFHPHKQEHSMKECPVQHQHQFFEQGDRICFSLRPLP
                     VCHSKCAATEKISKYFDVHCFEKDSTQAKKYKSEIGRGYTPDFKSFAPHKTYKFNYPK
                     SCVYKAY"
     polyA_site      6504
                     /note="16 A residues"
BASE COUNT     1748 a   1847 c   1303 g   1606 t
ORIGIN


  Query Match             6.9%;  Score 38;  DB 3;  Length 6504;
  Best Local Similarity   49.5%;  Pred. No. 0.48;
  Matches   98;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     289 ATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGC 348
           ||| ||||| |||        |  |  ||||| || ||||| ||   || |    || | |
Db    6265 ATCAGCAAATACTTCGATGTCCACTGCTTCGAGAAGGATTCCACCCAGGCTAAGAAGTAC 6324

Qy     349 CAGTCAGGCATCTGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATAT 408
           | || |  ||  |   | |    ||      | | |  ||  ||| ||     | |   |
Db    6325 AAATCCGAGATTGGCCGCGGCTACACTCCGGACTTCAAGAGCTTCGCCCCACACAAGACT 6384

Qy     409 TCAGTAACCAAGTCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGA 468
           |       ||| |   | || || | ||| |      |  |  | || | | ||| || |
Db    6385 TACAAGTTCAACTACCCGAAGAGCTGTGTCTACAAGGCATACTAGGAAACGACATTTGCA 6444

Qy     469 GATGGCAATATTGTGAGA 486
           |||  || | ||||  ||
Db    6445 GATCCCATTTTTGTATGA 6462


RESULT   6
MQSVITE
LOCUS       MQSVITE      8780 bp    DNA             INV       15-MAY-1996
DEFINITION  Aedes aegypti vitellogenin (VgA1) gene, complete cds.
ACCESSION   L41842
NID         g790834
KEYWORDS    vitellogenin.
SOURCE      Aedes aegypti (clone: 403-1) DNA.
  ORGANISM  Aedes aegypti
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Aedes.
REFERENCE   1  (bases 1 to 8780)
  AUTHORS   Romans,P., Tu,Z., Ke,Z. and Hagedorn,H.H.
  TITLE     Analysis of a vitellogenin gene of the mosquito, Aedes aegypti and
            comparisons to vitellogenins from other organisms
  JOURNAL   Insect Biochem. Mol. Biol. 25 (8), 939-958 (1995)
  MEDLINE   96035842
FEATURES             Location/Qualifiers
     source          1. .8780
                     /organism="Aedes aegypti"
                     /db_xref="taxon:7159"
                     /clone="403-1"
     exon            2016. .2121
                     /gene="VgA1"
                     /evidence=experimental
```

```
     gene            join(2016. .2121,2192. .7405,7463. .>8722)
     mRNA            /gene="VgA1"
                     join(2016. .2121,2192. .7405,7463. .>8722)
     CDS             /gene="VgA1"
                     join(2091. .2121,2192. .7405,7463. .8664)
                     /gene="VgA1"
                     /note="putative"
                     /codon_start=1
                     /product="vitellogenin"
                     /db_xref="PID:g790835"
                     /translation="MLAKLLLLALAGLTAAYQYENSFKGYNPGYKGYDAGYKGYGYDA
                     GYKGYGYDAGYKYNNQGYSYKNGFEYGYQNAYQAAFYKHRPNVTEFEFSSWMPNYEYV
                     YNVTSKTMTALAELDDQWTGVFTRAYLVIRPKSRDYVVAYVKQPEYAVFNERLPHGYA
                     TKFYHDMFKFQPMPMSSKPFGIRYHKGAIKGLYVEKTIPNNEVNILKAWISQLQVDTR
                     GANLMHSSKPIHPSKNEWNGHYKVMEPLVTGECETHYDVNLIPAYMIQAHKQWVPQGQ
                     LRGEDDQFIQVTKTQNFDRCDQRMGYHFGFTGYSDFRPNTNQMGNVASKSLVSYMYLT
                     GNWYNFTIQSSSMINKVAIAPSLVNKEPALVYAQVNMTLNDVHPYDKVPMGPAEDLKV
                     FVDLVYSYNMPSDKKNYVRPGNETSSSSSSSSSSSSSSESSSSSSESVENPKISPVE
                     QYKPLLDKVEKRGNRYRRDLNAIKEKKYYEAYKMDQYRLHRLNDTSSDSSSSDSSSSS
                     SSESKEHRNGTSSYSSSSSSSSSSSSSSESSSYSSSSSSSSESYSISSEEYYYQPTPAN
                     FSYAPEAPFLPFFTGYKGYNIFYARNVDAIRSVGKLVEEIASDLENPSNLPKANTMSK
                     FNILTRAIRAMGYEDIYELAQKYFVSQKERQVAQFSDKKFSKRVDAWVTLRDAVAEAG
                     TPSAFKLIFDFIKEKKLRGYEAATVIASLAQSIRYPTEHLLHEFFLLVTSDVVLHQEY
                     LNATALFAYSNFVNQAHVSNRSAYNYYPVFSFGRLADADYKIIEHKIVPWFAHQLREA
                     VNEGDSVKIQVYIRSLGNLGHPQILSVFEPYLEGTIQITDFQRLAIMVALDNLVIYYP
                     SLARSVLYRAYQNTADVHEVRCAAVHLLMRTDPPADMLQRMAEFTHHDPRLYVRAAVK
                     SAIETAALADDYDEDSKLALNAKAAINFLNPEDVSIQYSFNHIRDYALENLELSYRLH
                     YGEIASNDHRYPSGLFYHLRQNFGGFKKYTSFYYLVSSMEAFFDIFKKQYNTKYFADY
                     YKSADYSTNYYNFDKYSKYYKQYYYSKDSEYYQKFYGQKKDYYNDKEPFKFTAPRIAK
                     LLNIDAEEAEQLEGQLLFKLFNGYFFTAFDNQTIENLPHKMRHLFENLEDGYAFDVTK
                     FYQQQDVVLAWPLATGFPFIYTLKAPTVFKFEVDASAKTHPQVYKMPAGHPETENDDF
                     FYMPQSINGSVDVNLLYHRMVDAKVGFVTPFDHQRYIAGYQKKLHGYLPFNVELGLDF
                     VKDEYEFEFKFLEPKDDHLLFHMSSWPYTGYKDITDMRPIAENPNAKIVHDDNQSTKT
                     MEHTFGQDMTGVALRFHAKYDFDLINFQQFWSLIQKNDFVSAVNYPFAYQPYEYHQFN
                     LFYDSQRTHAKSFKFFAYQKFGAPSFEETGPKHPANRHSYSGNYYESNYAQPFVYSPG
                     SQRRYEQFFRNAASGIRNSFVRYYDFGFEFYAPQYKSEFTFTTAFADSPVDKTSRQLY
                     YFYASPMFPSQSYFKDIPFSGKQFQFCATATSEFPRVPYLKFSDFDKYYGDASQYFDF
                     LYGESCQGGAHIAVKGKQKQTGKCREYLRFSDVAKACKEQMANGYYQFEECQQAIDQA
                     YYYDFYDYAIEYKDVGSVAKNLTNKFYNYFQYAFYPYFESNFFYHGKSNYIKAEFEFA
                     PYGDYYNASFFGPSYAFQVQNYPVFNDYSTYFPYFFKYTFFPRYQPYYMHRLPSHKPR
                     NRPYYELSNYEQFAIFDRKPQYPSCSFSNDNFYTFDNKKYFYDMGECWHAVMYTVKPD
                     YDFYAQQSHFYNSDFEYKYKNGFEEYEQFAALARRGSDNQLYFKFLFGDNYIEVFPNN
                     GGVPFVKYNGRPYDISKSNIAHFEYKEGYPSFPFFYAFAYPNKDLEVSFFGGKLKFAT
                     DGYRARFFSDYSFYNNFVGLCGTNNGEYFDEFVTADQCYMRKPEFFAASYAITGQNCT
                     GPAKAFNYAYQQKAKQECVKREVYYGDIIYNQEYYHPRYRYYNHNVEESSSSSSSSSS
                     DSSSSSSSSEFSSLGRSGSSSSSSSSSEEQKEFHPHKQEHSMKECPVQHQHQFFEQGDR
                     ICFSLRPLPVCHSKCAATEKISKYFDVHCFEKDSTQAKKYKSEIGRGYTPDFKSFAPH
                     KTYKFNYPKSCVYKAY"
     intron          2122. .2191
                     /gene="VgA1"
     exon            2192. .7405
                     /gene="VgA1"
     intron          7406. .7462
                     /gene="VgA1"
     exon            7463. .>8722
                     /gene="VgA1"
     polyA_signal    8717. .8722
                     /gene="VgA1"
BASE COUNT     2468 a    2274 c    1733 g    2305 t
ORIGIN
```

```
  Query Match              6.9%;  Score 38;  DB 3;  Length 8780;
  Best Local Similarity   49.5%;  Pred. No. 0.5;
  Matches   98;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     289 ATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGC 348
           ||| ||||| |||       |  |  ||||| || ||||| ||   || |    || | |
Db    8500 ATCAGCAAATACTTCGATGTCCACTGCTTCGAGAAGGATTCCACCCAGGCTAAGAAGTAC 8559

Qy     349 CAGTCAGGCATCTGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATAT 408
           | || |  ||  |   | |    ||     | | |  ||  ||| ||     | |   |
Db    8560 AAATCCGAGATTGGCCGCGGCTACACTCCGGACTTCAAGAGCTTCGCCCCACACAAGACT 8619

Qy     409 TCAGTAACCAAGTCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGA 468
           |       ||| |   | || || | ||| |       |  |  | || | | ||| || |
Db    8620 TACAAGTTCAACTACCCGAAGAGCTGTGTCTACAAGGCATACTAGGAAACGACATTTGCA 8679

Qy     469 GATGGCAATATTGTGAGA 486
           ||| || | ||||  ||
Db    8680 GATCCCATTTTTGTATGA 8697
```

```
RESULT   7
SC2E1/c
LOCUS       SC2E1       38962 bp    DNA             BCT       04-JUN-1998
DEFINITION  Streptomyces coelicolor cosmid 2E1.
ACCESSION   AL023797
NID         g3191976
KEYWORDS    30S ribosomal protein S16; 30S ribosomal protein S2; 50S ribosomal
            protein L19; cell division protein FtsH homolog; elongation factor
            Ts; ffh; ftsH; glnB; glnD; nitrogen regulatory protein pII;
            prophage; protein pII uridylyltransferase; pyrH; RNA polymerase
            sigma factor WhiG; rplS; rpsB; rpsP; signal peptidase I; signal
            recognition particle protein; sipl; sip2; sip3; sip4; trmD; tRNA
            (guanine-N1)-methyltransferase; tsf; uridylate kinase; whiG.
SOURCE      Streptomyces coelicolor.
  ORGANISM  Streptomyces coelicolor
            Eubacteria; Firmicutes; Actinomycetes; Streptomycetes;
            Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 38962)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 38962)
  AUTHORS   Murphy,L. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 38962)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
```

```
          sequenced clone.  It may be shorter because we only sequence
          overlapping sections once, or longer, because we arrange for a
          small overlap between neighbouring submissions. Cosmid 2E1 lies
          between 7A1 and 6A9 on the AseI-B genomic restriction fragment.
FEATURES             Location/Qualifiers
     source          1. .38962
                     /organism="Streptomyces coelicolor"
                     /strain="A3(2)"
                     /db_xref="taxon:1902"
                     /clone="cosmid 2E1"
     gene            1. .166
                     /gene="glnB"
     CDS             <1. .166
                     /gene="glnB"
                     /note="SC2E1.01, glnB, nitrogen regulatory protein pII,
                     partial CDS, len: >54 aa; highly similar to many e.g.
                     GLNB_AZOBR (112 aa), fasta scores; opt: 198 z-score: 381.9
                     E(): 4.9e-14, 50.0% identity in 54 aa overlap"
                     /codon_start=2
                     /transl_table=11
                     /product="nitrogen regulatory protein pII"
                     /db_xref="PID:e1296621"
                     /db_xref="PID:g3191977"
                     /translation="IRIEVLVEDDDAEQLIDVVVKAARTGKIGDGKVWAVPVDTAVRV
                     RTGERGPDAL"
     gene            231. .2738
                     /gene="glnD"
     CDS             231. .2738
                     /gene="glnD"
                     /note="SC2E1.02, probable glnD, protein pII
                     uridylyltransferase, len: 835 aa; similar to many e.g.
                     GLND_ECOLI (890 aa), fasta scores; opt: 403 z-score: 540.4
                     E(): 7.3e-23, 26.2% identity in 851 aa overlap. Contains 7
                     degenerate repeats of the sequence PSS"
                     /codon_start=1
                     /transl_table=11
                     /product="putative protein pII uridylyltransferase"
                     /db_xref="PID:e1296622"
                     /db_xref="PID:g3191978"
                     /translation="MTDEAEDSGPGGYAAARLRLLTEGARSGPPRRRALAELTDGWLA
                     GLFGAATEEHTGISLVAVGGYGRGELSPRSDLDLLLLHDGRDDKAVAALADRLWYPVW
                     DLGIDLDHSVRTPQQARKTAGQDLKVHLGLLDARHLAGDLGLTASLRTAVLADWRNQA
                     PKRLPELRDLCAERAERQGELQFLLEPDLKEARGGLRDATALRAVAASWLADAPREGL
                     AEARRRLLDVRDALHLATGRATDRLALQEQDQVAAELGLLDADALLRQVYEAARVISY
                     AGDVTWREVGRVLRSRSVRPRLRAMMNGRNGGKPVAERSPLAEGVVEQDGEAVLARTA
                     RPERDPALPLRAAAAAAQAGLPLSRHAVRRLAATARPLPTPWPAEAREQLVTLLGSGR
                     PTVQVWEALEAEGLVTRLLPDWERVRCRPQRNAVHLWTVDRHLIETAVRAAGFTRRVH
                     RPDLLLIAALLHDIGKGWPGDHSVAGETIARDVAARIGFDGADTAVLATLVRHHLLLV
                     ETATRRDLDDPATVRAVAQAVGTEHTLELLHALTEADALATGPAAWSSWRGSLVADLV
                     KRVSGVLAGEPQPEAESAAPTAEQERLAVEAFRTGGPVLALRAQTEPPADSAPAPSSP
                     SSPSFPSPLSSPSSPSSADGPEPLGVELLIAVPDQAGVLPAVAGVLAMHRLTVRTAEL
                     RSVPLPDGVEGSVLLLDWRVAAQYGSLPQAARLRADLVRALDGTLDIAARLAERDAAH
                     PRRRGVEPPPPRVTVAPAASRLATVIEVRAQDAPGLLFRLGRALEAAGVRVRSAHVST
                     LGANAVDAFYVTRGEGTPLPGDEAASVARGLEESLRT"
     repeat_region   2091. .2153
                     /note="7 degenerate 9bp repeats"
     RBS             2719. .2723
                     /gene="glnD"
                     /note="possible RBS upstream of ffh"
     gene            2735. .4387
                     /gene="ffh"
     CDS             2735. .4387
                     /gene="ffh"
                     /note="SC2E1.03, ffh, signal recognition particle protein,
                     len: 550 aa; highly similar to mnay e.g. SR54_ECOLI (453
                     aa), fatsa scores; opt: 1465 z-score: 1282.8 E(): 0, 50.0%
                     identity in 458 aa overlap. Contains PS00017
                     ATP/GTP-binding site motif A (P-loop), PS00300 SRP54-type
                     proteins GTP-binding domain signature and Pfam match to
                     entry SRP54 PF00448, SRP54-type proteins, score 66.57.
                     Similarities suggest possible start site at aa 35"
                     /codon_start=1
                     /transl_table=11
                     /product="signal recognition particle protein"
                     /db_xref="PID:e1296623"
                     /db_xref="PID:g3191979"
                     /translation="MTVRSVQRAGTLCLPGRIPWRTTHTNADARTYAAVFDTLSDRLS
                     ATFKSLRGKGRLTEADIDATAREIRIALLEADVALPVVRAFIKKVKERSLGAEVSKAL
                     NPAQQVLKIVNEELVGILGGETRRLRFAKQPPTVIMLAGLQGAGKTTLAGKLGHWLKE
                     QGHSPLLVACDLQRPNAVNQLSVVAERAGVAVYAPEPGNGVGDPVKVAKDSIEFAKAK
                     VHDLVIVDTAGRLGIDQELMQQAADIRDAVSPDEILFVVDAMIGQDAVNTAEAFRDGV
                     GFDGVVLSKLDGDARGGAALSIASVTGKPIMFASNGEKLDDFDAFHPDRMASRILDMG
                     DLLTLIEQAEKTFSQEEAEKMASKLASKKGQDFTLDDFLAQMEQVRKMGSISKLLGML
                     PGMGQMKDQINNLDERDVDRTAAIIKSMTPGERQEPTIINGSRRARIAKGSGVEVSAV
                     KNLVERFFEARKMMSRMAQGGGMPGMPGMPGMGGGPGRQKKQQKKAKGKQRSGNPMKR
                     KQQEQEEAARRAAAAQSGGALGLPQQGGQDFELPDEFKKFMG"
     misc_feature    3155. .3178
                     /gene="ffh"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     misc_feature    3566. .3697
                     /gene="ffh"
                     /note="Pfam match to entry SRP54 PF00448, SRP54-type
                     proteins, score 66.57"
     misc_feature    3647. .3688
                     /gene="ffh"
                     /note="PS00300 SRP54-type proteins GTP-binding domain
                     signature"
     stem_loop       4405. .4453
                     /note="hairpin loop with 22/23 bp stem"
     gene            4504. .6450
                     /gene="ftsH"
     CDS             4504. .6450
                     /gene="ftsH"
                     /note="SC2E1.04, ftsH, cell division protein FtsH homolog
                     (zinc metalloprotease, integral membrane protein), len:
                     648 aa; similar to many e.g. FTSH_LACLA (695 aa), fasta
                     scores; opt: 1489 z-score: 1195.5 E(): 0, 41.7% identity
                     in 640 aa overlap. Contains PS00017 ATP/GTP-binding site
                     motif A (P-loop), PS00674 AAA-protein family signature and
                     Pfam matches to entry AAA PF00004, ATPases associated with
                     various cellular activities (AAA) and to entry zn-protease
                     PF00099, Zinc-binding metalloprotease domain, score 22.07,
                     score 357.96. Proline-rich N-terminus is not present in
                     other ftsH homologs"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein FtsH homolog"
                     /db_xref="PID:e1296624"
                     /db_xref="PID:g3191980"
                     /translation="MTNPSPPRKAPEPPWRTEGTPDEPPKPPPGGRRMRGGWWNLILA
                     ALIVYLIANLVLSFFNEGDEPTISYTEFSKQVDEGNVSKIYAKGDAIQGQLKKARDNP
                     EGDGTYTKFTTERPTFADDQLWADLTKNRVTVTAEPVVQHRSFLANLLIALAPMLILV
                     VLWIFIARRMRGALGGGAGGMLGRKAPPKPVELEAGKPRTTFADVAGIDEVEGELSDV
                     VDFLKNPDAYRRMGAKMPRGVLLTGPPGTGKTLLARAVAGEAGVPFFSASASEFIEMI
                     VGVGASRVRELFAEARKVAPSIIFIDEIDTIGRARGGGSGTGGHDEREQTLNQILTEM
                     DGFSGSEGVIVIAATNRADILDAALTRPGRFDRVVSVSPPDRGGREAILEIHTREIPL
                     APDIDLAQVARTTPGMTGAELANLANEAALLAVKRKQERVTQANLSEALEKVQLGAER
                     PLVMPEEERRRTAYHESGHALLGMLQPGADPVRKITIVPRGRALGVTLSTPDADKYAY
                     TEEYLRGRIIGALGGMAAEHVVYGMITTGSESDLEQVTNIARGMVARWGMSERVGRLS
                     ALPGDAQQAYGLAAAPQTLDAIDGEMRRVVDSCYEEAVRKLRDHRGQLDALAESLLAS
                     ETLDEADAYRIAGITRLTKDDPEA"
     misc_feature    5215. .5781
                     /gene="ftsH"
                     /note="Pfam match to entry AAA PF00004, ATPases associated
                     with various cellular activities (AAA), score 357.96"
     misc_feature    5230. .5253
                     /gene="ftsH"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     misc_feature    5530. .5586
                     /gene="ftsH"
                     /note="PS00674 AAA-protein family signature"
     misc_feature    5884. .5931
                     /gene="ftsH"
                     /note="Pfam match to entry zn-protease PF00099,
                     Zinc-binding metalloprotease domain, score 22.07"
     gene            complement(6468. .7085)
```

```
                /gene="SC2E1.05c"
CDS             complement(6468. .7085)
                /gene="SC2E1.05c"
                /note="SC2E1.05c, probable lipoprotein, len: 205 aa;
                contains N-terminal signal sequence and appropriately
                positioned PS00013 Prokaryotic membrane lipoprotein lipid
                attachment site"
                /codon_start=1
                /transl_table=11
                /product="putative lipoprotein"
                /db_xref="PID:e1296625"
                /db_xref="PID:g3191981"
                /translation="MPHRTTALRLAAPVAVLALALTACGGGDDDSGTDAKGGATTEAT
                TQASPSPSQPASGGDMAQGESASGKVAEGPGEVTYDVLAQKVEVGTEAEAAKAVADAA
                DAKGKVLATAHIKYTHQGGPALTEGSDVHDGTTVFADGQRGGVLIGAAEDAPGCEDPY
                DAESWKQGESHVFCETYVIPAGATSVEVHWSEEDGEPYVWKFPQK"
misc_feature    complement(7014. .7046)
                /gene="SC2E1.05c"
                /note="PS00013 Prokaryotic membrane lipoprotein lipid
                attachment site"
RBS             complement(7092. .7096)
                /note="possible RBS upstream of Sc2E1.05c"
gene            complement(7236. .8123)
                /gene="SC2E1.06c"
CDS             complement(7236. .8123)
                /gene="SC2E1.06c"
                /note="SC2E1.06c, unknown, len: 295 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.06c"
                /db_xref="PID:e1296626"
                /db_xref="PID:g3191982"
                /translation="MTPTLVRQHSSHAQATPHVDPRARARDWSEIQERMLVPLYEAVY
                ERLDVGGDTRMLGLNCGSGLALLMAASRGAAVTGVDSRSPELLDLARRRLLSGTSGDP
                GADSTVSTRTRGDAPRIVQGTPGDAADPTAPAFTLVTAFEPGGCRAGDSETLGDLLAR
                ATPLAARGAAVVLAGWGPPERCATTSVLRVAAKLTDPTHGTGGWRPALRDDLEEVAQR
                AGLRPDGSGRVACPFGYADTGSAVRGLLSTGMFDAAVAATDRDQVDKELTEALHPYQR
                PDGTVWMPNVFRYLIARVR"
gene            8793. .9389
                /gene="SC2E1.07"
CDS             8793. .9389
                /gene="SC2E1.07"
                /note="SC2E1.07, unknown, len: 198 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.07"
                /db_xref="PID:e1296627"
                /db_xref="PID:g3191983"
                /translation="MSTTPPVLAAELAQAWADIQRHHTELPDLAAPESLIGESSSACG
                HELSFERLLHEAVHGIAAARGVRDTSRAGRYHNRRFLAIAEELGLDHPEEPHPSSGFS
                LVTLTPEAKRRYRPTIERLQRALKAHTVATATDTARSFRGPAARHGSSGGGVRVKAVC
                DCGRNVRVVPSVLAQAPIMCGGCGKPFRIPEVVGAGMS"
RBS             9629. .9633
                /note="possible RBS upstream of rpsP"
gene            9645. .10064
                /gene="rpsP"
CDS             9645. .10064
                /gene="rpsP"
                /note="SC2E1.08, rpsP, 30S ribosomal protein S16, highly
                similar to many e.g. RS16_MYCTU (162 aa), fatsa scores;
                opt:480 z-score: 771.3 E(): 0, 56.3% identity in 144 aa
                overlap. Contains PS00732 Ribosomal protein S16 signature"
                /codon_start=1
                /transl_table=11
                /product="30S ribosomal protein S16"
                /db_xref="PID:e1296628"
                /db_xref="PID:g3191984"
                /translation="MAVKIKLKRLGKIRSPHYRIVVADSRTRRDGRAIEEIGKYHPTY
                NPSVMEVDAERVAYWLGVGAQPTEPVLAILKKTGDWQKFKGEPAPAPLLQPAEKAARP
                SFEAIGGEDEGKGEAITQKKKADKKDEAAAESSASEA"
misc_feature    9651. .9680
                /gene="rpsP"
                /note="PS00732 Ribosomal protein S16 signature"

RBS             10056. .10059
                /gene="rpsP"
                /note="possible RBS upstream of SC2E1.09"
gene            10067. .10306
                /gene="SC2E1.09"
CDS             10067. .10306
                /gene="SC2E1.09"
                /note="SC2E1.09, unknown, len: 79 aa; highly similar to
                small hypothetical proteins from several organisms, e.g.
                YX40_MYCTU MTCY274.40c (80 aa), fasta scores; opt: 356
                z-score:628.4 E(): 9.1e-28, 71.6% identity in 74 aa
                overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.09"
                /db_xref="PID:e1296629"
                /db_xref="PID:g3191985"
                /translation="MLEEALEHLVKGIVDNPDDVQVASRNLRRGRVLEVRVHPDDLGK
                VIGRNGRTARALRTVVGAIGGRGVRVDLVDVDHVR"
RBS             10403. .10407
                /note="possible RBS upstream of SC2E1.10"
gene            10418. .10984
                /gene="SC2E1.10"
CDS             10418. .10984
                /gene="SC2E1.10"
                /note="SC2E1.10, unknown, len: 188 aa; similar to
                hypothetical proteins from many organisms e.g. YX38_MYCTU
                MTCY274.38c (176 aa), fasta scores; opt: 479 z-score:
                693.1 E(): 2.3e-31, 48.3% identity in 176 aa overlap. Also
                some similarity to RIMM_HAEIN 16S rRNA processing protein
                rimM (178 aa), fasta scores; opt: 210 z-score: 174.2 E():
                0.018, 27.6% identity in 170 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.10"
                /db_xref="PID:e1296630"
                /db_xref="PID:g3191986"
                /translation="MQLVVARIGRAHGIKGEVTVEVRTDEPELRLGPGAVLATDPAST
                GPLTIESGRVHSGRLLLRFAGVHDRTGAEALRNTLLIADVDPDERPEDEDEYYDHQLI
                DLDVVTEDGTEVGRITEISHLPTQDLFVVERPDGSEVYVPFVSEIVTGIDLDAQRAVI
                DPPPGLIDDRAEIASARDAGAEDAGDEA"
RBS             10976. .10979
                /gene="SC2E1.10"
                /note="possible RBS upstream of trmD"
gene            10984. .11817
                /gene="trmD"
CDS             10984. .11817
                /gene="trmD"
                /note="SC2E1.11, trmD, probable tRNA
                (guanine-N1)-methyltransferase, len: 277 aa; similar to
                many e.g. TRMD_ECOLI (EC2.1.1.31) (255 aa), fasta scores;
                opt: 437 z-score: 660.6E(): 1.5e-29, 41.2% identity in 238
                aa overlap"
                /codon_start=1
                /transl_table=11
                /product="tRNA (guanine-N1)-methyltransferase"
                /db_xref="PID:e1296631"
                /db_xref="PID:g3191987"
                /translation="MRLDVVTIFPEYLEPLNVSLVGKARARGQLGVHVHDLRDWTYDR
                HNTVDDTPYGGGPGMVMKTEPWGDALDSVLADGYETGCGEPALVVPTPSGRPFTQELA
                VHLSERPWLIFTPARYEGIDRRVVDEYATRMPVYEVSIGDYVLAGGEAAVLVVTEAVA
                RLLPGVLGNAESHRDDSFAPGAMANLLEGPVHTKPPQWRGRGIPDVLLSGHHGKIARW
                RRDEALRRTTANRPDLIERCDPAAFDKKDREMLSILGWQPDPDGEPYGRFWRRIPGME
                E"
RBS             11939. .11942
                /note="possible RBS upstream of rplS"
gene            11954. .12304
                /gene="rplS"
CDS             11954. .12304
                /gene="rplS"
                /note="SC2E1.12, rplS, 50S ribosomal protein L19, len:
                116aa; highly similar to many e.g. RL19_MYCTU (113 aa),
                fastascores; opt: 571 z-score: 1138.4 E(): 0, 77.3%
```

identity in110 aa overlap. Contains PS01015 Ribosomal protein L19 signature"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L19"
/db_xref="PID:e1296632"
/db_xref="PID:g3191988"
/translation="MSHLLDSVDAASLRSDVPAFRPGDTVNVHVRVIEGNRSRVQQFK
GVVIRRQGAGVRETFTVRKVSFSVGVERTFPVHTPIVEKIELVTRGDVRRAKLYYLRE
LRGKAAKIKEKRDS"

misc_feature 12215. .12256
/gene="rplS"
/note="PS01015 Ribosomal protein L19 signature"

gene 12350. .13129
/gene="sip1"

CDS 12350. .13129
/gene="sip1"
/note="SC2E1.13, sip1, probable signal peptidase I, len: 259 aa; similar to many e.g. LEP_HAEIN signal peptidase I (EC 3.4.21.89) (349 aa), fasta scores; opt: 162 z-score: 381.2 E(): 5.3e-14, 32.5% identity in 243 aa overlap, though with large internal deletions. Contains PS00761 Signal peptidases I signature 3, Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 51.12 and probable transmembrane domain. Also similar to downstream genes sip2 (E(): 9.7e-26, 44.9% identity in 274 aa overlap), sip3 (E(): 4.8e-16, 35.6% identity in 264 aa overlap) and sip4 (E(): 8.7e-08, 37.0% identity in 219 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296633"
/db_xref="PID:g3191989"
/translation="MDTETQHTERDRSSRPSDSEQPSEPEGPEERSRSAFTGRIGDWL
PGGRITVTLLILLLFLLLVSTFVLQPFQIPSGSMERGLRIGDRVLVNKLAYRFDGRPR
RGDIVVFDGTGSFGHGDYIKRVVGVGGDHVVCCDEEGRVRVNGQPVDESAFLYPGDRP
STVPFDVVVPDGTLFVLGDHRSDSSDSRDHLGSPGGGMVPLDDVIGRADWIVWPFGHA
TRLDRPDAYARVPDAEAGAGAASGAGAAPDPGAAPDGDADG"

stem_loop 12369. .12455
/gene="sip1"
/note="hairpin loop with 34/41 bp stem"

misc_feature 12509. .12799
/gene="sip1"
/note="Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 51.12"

misc_feature 12869. .12910
/gene="sip1"
/note="PS00761 Signal peptidases I signature 3"

gene 13122. .14462
/gene="sip2"

CDS 13122. .14462
/gene="sip2"
/note="SC2E1.14, sip2, probable signal peptidase I, len: 446 aa; similar to many e.g. LEP_HAEIN signal peptidase I (EC 3.4.21.89) (349 aa), fasta scores; opt: 151, z-score: 317.3 E(): 1.9e-10, 30.1% identity in 286 aa overlap. Contains PS00761 Signal peptidases I signature 3, Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 42.41 and probable transmembrane domain. Also similar to surrounding genes sip1 (E(): 9.7e-24, 44.9% identity in 274 aa overlap), sip3 (E(): 1.8e-28, 56.3% identity in 245 aa overlap), and sip4 (E(): 1.1e-09, 35.9% identity in 223 aa overlap). Note that after aa 325 the ORF overlaps the downstream sip3 gene SC2E1.15, and positional base preference suggests that translation is in the other frame. No error in the cosmid sequence can be found to account for this"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296634"
/db_xref="PID:g3191990"
/translation="MGSRGKPRGAPSSPAENLLPTGSRRTAAPSGGRSRAERRKLQRK
VKRRRRRGAVKEIPLLVGVAVLIALVLKTFLVQAFVIPSGSMEQTIRIGDRVLVDKLT
PWFGSEPQRGDVVVFRDPGGWLQGEQTTKKDDPVVVKQVKEGLAFIGLLPSDDEKDLI
KRVVGVGGDHVKCCDKQGRVTVNGVPLTEDYLYPGDRPSRTPFDVTVPEGRLWVMGDH
RSNSADSRAHQETDFGTVSQDEVVGRAMVIAWPFGHWTTLDEPKTYASVSDAASGSTA
APELSHRVAPYDPNAMMELPTPAELPLVMGVVGLRRIRGRRRQRVRSWRGGCGGWRTV
RSRRRGEPRTPRGEVRSGRGRRPRLRTRLRNRGRQGDERTARAGRTQRRDGGRGSRRD
ATRDPARGQEAALLLEGAADPGRHRAGARPADQDVPGAGVLDPVLLDGEHPSDR"

misc_feature 13236. .13484
/gene="sip2"
/note="Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 42.41"

misc_feature 13755. .13796
/gene="sip2"
/note="PS00761 Signal peptidases I signature 3"

misc_feature 13859. .15213
/note="97.8% match to EMBL:SLTK24SIP Y09474 Streptomyces lividans sip gene, from 5 to 1359"

RBS 14085. .14089
/gene="sip2"
/note="possible RBS upstream of sip3"

gene 14095. .15105
/gene="sip3"

CDS 14095. .15105
/gene="sip3"
/note="SC2E1.15, sip3, probable signal peptidase I, len: 336 aa; identical to, but longer than, S. lividans sip TR:E1228248 (EMBL:Y09474) signal peptidase I (291 aa), fasta scores; opt: 1997 z-score: 2287.4 E(): 0, 99.7% identity in 291 aa overlap, and similar to many others e.g. LEP_PHOLA signal peptidase I (EC 3.4.21.89) (203 aa), fasta scores; opt:222 z-score: 404.9 E(): 2.6e-15, 37.9% identity in 232 aa overlap. Contains PS00761 Signal peptidases I signature 3, 2x Pfam match to entry signal_pept_I PF00461, Signal peptidases I, scores 48.83 and 46.79 and probable transmembrane domain. Also similar to surrounding genes sip1 (E(): 7.6e-20, 35.6% identity in 264 aa overlap), sip2 (E(): 0, 56.3% identity in 245 aa overlap) and sip4 (E(): 6.6e-15, 36.3% identity in 237 aa overlap). Note that a more upstream start was chosen due to positional base preference"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296635"
/db_xref="PID:g3191991"
/translation="MGDVAVGARSGHDGEENRGRPEERSGPAAVGAPDSGHGSGTEDG
RVTNGQRGQDGHNGGTEDEGVGGTPPATPPAAKKQRSFWKELPILVGIALVLALLIKT
FLVQAFSIPSSSMENTLQIGDRVLVDKLTPWFGSEPERGEVVVFHDPADWLAGEPTPD
PNALQTVLSWIGLMPSAEEKDLIKRVIGVAGDTVECNKTGPLKVNGKALNEPYVYPGN
TPCSDDDQGGRFKVTVPEGKIWVMGDHRQNSRDSRYNQSDKNGGMVPVDEVVGRAIVV
AWPMNRWGTLPVPDTFDQDGLQARSSAAAALSVAPQGLAVAGVVPFVWWRRRRTAPAE
TR"

misc_feature 14302. .14547
/gene="sip3"
/note="Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 48.83"

misc_feature 14632. .14907
/gene="sip3"
/note="Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 46.79"

misc_feature 14809. .14850
/gene="sip3"
/note="PS00761 Signal peptidases I signature 3"

RBS 15197. .15202
/note="possible RBS upstream of sip4"

gene 15210. .15986
/gene="sip4"

CDS 15210. .15986
/gene="sip4"
/note="SC2E1.16, sip4, probable signal peptidase I, len: 258 aa; similar to many e.g. LEPS_BACSU signal peptidase I (EC 3.4.21.89) (184 aa), fasta scores; opt: 258 z-score:

337.6 E(): 1.4e-11, 32.2% identity in 177 aa overlap. Contains Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 49.57 and probable transmembrane domain. Also similar to upstream genes sip1 (E(): 8.8e-10, 36.5% identity in 219 aa overlap), sip2 (E(): 3.8e-13, 35.9% identity in 223 aa overlap) and sip3 (E(): 8.1e-15, 35.4% identity in 237 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296636"
/db_xref="PID:g3191992"
/translation="MGGESTTRTVPRGGGANKGPVGSRTGQRLSGIAVALGLVLFLGG
FAWGAVVYRPYTVPTSSMTPTIDAGDRVLAQRIDGADVRRGDVVVFKDATWANAPMVK
RVVAVGGDTVSCCQQGKLKVNGKVIDEPYLPAGTPAEISDFQTVTVPEGRLFLLGDER
RNSVDSTAHLTDAAAGTVSRGAVDARVDAVAWPMDGMLERPTGFRTLGDLSSPGPLRT
VVVLVIAGAVLILGGAAYGPLAKRAAASRARKGAEPAGGR"

misc_feature 15255. .15605
/gene="sip4"
/note="Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 49.57"

gene 15988. .16506
/gene="SC2E1.17"

CDS 15988. .16506
/gene="SC2E1.17"
/note="SC2E1.17, unknown, len: 172 aa; similar to C-terminus of M. tuberculosis hypothetical protein TR:O06205 (EMBL:Z95387) MTCY1A10.24 (351 aa), fasta scores; opt: 345 z-score: 375.2 E(): 1.2e-13, 44.7% identity in 161 aa overlap. Contains PS00893 mutT domain signature and Pfam match to entry mutT PF00293, Bacterial mutT protein, score 37.78"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.17"
/db_xref="PID:e1296637"
/db_xref="PID:g3191993"
/translation="MTPAGQDTYEGGLRRVARVVLLDPEDRILLLHGHEPDDPADDWW
FTPGGGVEGDETRAEAARRELLEETGITDVELGPVLWRRRCSFPFAGRRWDQDEWYYL
ARTARTATEAVGPGLTELERRSVAGARWWTCEELTRARETVYPTRLAELLTTLLDEGP
PAGPVTLDTEIV"

misc_feature 16069. .16200
/gene="SC2E1.17"
/note="Pfam match to entry mutT PF00293, Bacterial mutT protein, score 37.78"

misc_feature 16132. .16191
/gene="SC2E1.17"
/note="PS00893 mutT domain signature"

gene 16563. .16871
/gene="SC2E1.18"

CDS 16563. .16871
/gene="SC2E1.18"
/note="SC2E1.18, unknown, len: 102 aa; highly similar to M. tuberculosis hypothetical protein YX32_MYCTU MTCY274.32c (101 aa), fasta scores; opt: 515 z-score: 1085.2 E(): 0, 75.0% identity in 100 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.18"
/db_xref="PID:e1296638"
/db_xref="PID:g3191994"
/translation="MSAEDLEKYETEMELKLYREYRDVVGLFKYVIETERRFYLTNDY
EMQVHSVQGEVFFEVSMADAWVWDMYRPARFVKQVRVLTFKDVNIEELNKSDLELPGG
"

gene 16930. .17322
/gene="SC2E1.19"

CDS 16930. .17322
/gene="SC2E1.19"
/note="SC2E1.19, unknown, len: 130 aa; similar to M. tuberculosis hypothetical protein YX29_MYCTU MTCY274.29c (128 aa), fasta scores; opt: 253 z-score: 276.8 E(): 3.5e-08, 39.7% identity in 121 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.19"
/db_xref="PID:e1296639"
/db_xref="PID:g3191995"
/translation="MIVGAGGGADMNARGAMGRYGETLAARRLTGAGMTVLERNWRCG
RTGEIDIVARDGDVLVVCEVKTRRGGAFEHPMAAVTPDKAERLRRLAERWIQTHGGAP
PGGVRIDLVGVLLPQRGAPVVEHARGVA"

gene 17322. .18947
/gene="SC2E1.20"

CDS 17322. .18947
/gene="SC2E1.20"
/note="SC2E1.20, unknown ATP/GTP binding protein, len: 541aa; similar to hypothetical proteins from many organisms e.g. YIFB_ECOLI (516 aa), fasta scores; opt: 356 z-score: 935.9 E(): 0, 37.0% identity in 513 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.20"
/db_xref="PID:e1296640"
/db_xref="PID:g3191996"
/translation="MGFARTCSVALVGVEGVVVEVQADLEPGVAAFTLVGLPDKSLTE
SRDRVRAAVVNSGAQWPQKKLTVGLSPASVPKSGSGFDLAVAASVLGAAERIDPRVLA
DIVMIGELGLDGRVRPVRGILPAVLAAAEAGYEQVVVPECAAPEASLVPGVSVLGVRS
LRQLIAVLADEPVPEEEQDLQGRPDPLMAGLRMPGTGAATGMHGMGAAQHDHGHDLAD
VVGQESARTAVEVAAAGRHHLFLEGPPGAGKTMLAERLPAILPRLCRAESLEVTAVHS
VAGLLPPGKPLIDVAPYCAPHHSATMQALVGGGPGIARPGAVSLAHRGVLFLDETPEF
SSHALDALRQPLEAGHVVIARSAGVVRFPARFLMVLAANPCPCGRFSRTDELCECPSS
AIRRYQARLSGPLLDRVDLRVEVEVRVTRSQLAGDGPRGDSTRAVADRVRAARDRAAAR
LAGTPWRSNSEVPGRELRTRWRAAAGALDEAERSLERGVLTARGLDRVLRVAWTVADL
VGHDRPDARDVALALQLRTGVPRGVPMALGALA"

misc_feature 18048. .18071
/gene="SC2E1.20"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

gene 18944. .20092
/gene="SC2E1.21"

CDS 18944. .20092
/gene="SC2E1.21"
/note="SC2E1.21, unknown, len: 382 aa; N-terminus is Gly-rich, C-terminus is similar to the N-terminus of several hypothetical proteins e.g. SMF_ECOLI SMF protein (374 aa), fasta scores; opt: 402 z-score: 277.5 E(): 3.2e-08, 40.7% identity in 194 aa overlap. This ORF appears to be interrupted by the following prophage at around aa 360, and to continue as the last part of SC2E1.37. The similarity to the M. tuberculosis hypothetical protein YX27_MYCTU MTCY274.27c (389 aa) (fasta scores; opt: 603 z-score: 511.5 E(): 3e-21, 46.8% identity in 267 aa overlap) stops at that position and restarts in SC2E1.37. Alternative start site at aa 118 sugested by positional base preference"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.21"
/db_xref="PID:e1296641"
/db_xref="PID:g3191997"
/translation="MSGNTRPPEGAAPSGDTGPGTSGTLFPEAAVSFGGGSPAEGGAM
TPPGAGDGPDGGEARWAGPGPGDGDGAVTGAGVAARGRSAAGTCGTGGGGAGSGARGG
GPEDRELLGRVFLARVFEPGDEAGGRWVRERGAPEVVRRLREGGRALPGVSEKRWAGL
CARAGRADPGRDLAVARSAGARFVVPGTAEWPGQLDDLGDARPLGLWVRGGPSLRMWA
LRSVAVVGARACTEYGAHMAATLAAGLAERGWVVVSGGAYGIDGAAHRGALGAGGATA
AVLACGVDRPYPPGHTALITRIAEQGLVVGELPPGDHPTPSRFILRNRVIAALTRGTV
VVEAAHRSGSLVTARAARRLGRHVMGVPTPLSGGAALAPIRFRLATAT"

misc_feature 20021. .32836
/note="probable prophage: has low GC (approx 66-68%), and some ORF's with similarities to bacteriophage ORF's. The ORF SC2E1.21 appears to have been interrupted by this prophage and apparently continues as the last part of SC2E1.37"

gene 20271. .20582
/gene="SC2E1.22"

CDS 20271. .20582
/gene="SC2E1.22"

```
                /note="SC2E1.22, unknown prophage gene, len: 103 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.22"
                /db_xref="PID:e1296642"
                /db_xref="PID:g3191998"
                /translation="MTTQNSNHPCACGSYAFEVLIHENVGGDKVWQQKTTGCAATTQS
                TFAPGHDAKLKSLLIAAGVGGHPVRQTTRDTVVVKDALKVAADLGWRDLVGEAIAKGS
                S"
gene            complement(20579. .20869)
                /gene="SC2E1.23c"
CDS             complement(20579. .20869)
                /gene="SC2E1.23c"
                /note="SC2E1.23c, unknown prophage gene, len: 96 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.23c"
                /db_xref="PID:e1296643"
                /db_xref="PID:g3191999"
                /translation="MTTRYPALPQSYAITRDRDTVIPRYLVVTALRGGQMAGLSIRVY
                KIVGETGETEEVSAREWSVDQMPNLPLIDTRWPLCLCWKCNGNGTKGELRRR"
RBS             20912. .20916
                /note="possible RBS upstream of SC2E1.23"
misc_feature    20920. .21099
                /gene="SC2E1.24"
                /note="Pfam match to entry gntR PF00392, Bacterial
                regulatory proteins, gntR family, score 43.92"
gene            20920. .21114
                /gene="SC2E1.24"
CDS             20920. .21114
                /gene="SC2E1.24"
                /note="SC2E1.24, possible prophage transcriptional
                regulator, len:64 aa; similar to the N-terminus of several
                regulatory proteins e.g. Streptomyces phaeochromogenes
                plasmid pJV1 TR:Q54677 (EMBL:U23762) traR (245 aa), fasta
                scores; opt:137 z-score: 268.5 E(): 1e-07, 39.3% identity
                in 61 aa overlap. Contains Pfam match to entry gntR
                PF00392, Bacterial regulatory proteins, gntR family, score
                43.92"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /db_xref="PID:e1296644"
                /db_xref="PID:g3192000"
                /translation="MAEIISKRIADGTYPPGTRVPGIVELSAEFGIAASTAQKALAHL
                RGTGEVRTELGLGTFVADAK"
RBS             21200. .21204
                /note="possible RBS upstream of SC2E1.25"
gene            21209. .21910
                /gene="SC2E1.25"
CDS             21209. .21910
                /gene="SC2E1.25"
                /note="SC2E1.25, unknown prophage gene, len: 233 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.25"
                /db_xref="PID:e1296645"
                /db_xref="PID:g3192001"
                /translation="MATTDGDEPPEEREFLTGDTPFVRVREALERLAALNEGPRVEQR
                VDRVDQIQNVFIGQGQDTHGREVAERAATRLAEQRQKFHFDFLNHALKQAEWTFRLSV
                WFMTGGALVILAGAVLALVHAGNPDLSYLPLVTSLTGALITVGGGALALHSKRAMANL
                TKAAEDNENKIDIDHKLEVATTFIDRVQDSQAKDRLNSAAALKALGMDAQPETMVNRL
                LPDPQTKEIDPGEPT"
gene            complement(21969. .22166)
                /gene="SC2E1.26c"
CDS             complement(21969. .22166)
                /gene="SC2E1.26c"
                /note="SC2E1.26c, unknown prophage gene, len: 65 aa; very
                hydrophobic"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.26c"
                /db_xref="PID:e1296646"
                /db_xref="PID:g3192002"
                /translation="MNLPLGVKVLIVVICTLISVIVAMVAGFLSHASGSSVGQAILYG
                GGAFAGSLLLCLAVLSALGVL"
RBS             complement(22177. .22180)
                /note="possible RBS upstream of SC2E1.26c"
gene            complement(22505. .22867)
                /gene="SC2E1.27c"
CDS             complement(22505. .22867)
                /gene="SC2E1.27c"
                /note="SC2E1.27c, unknown prophage gene, len: 120 aa;
                contains hydrophobic region near centre"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.27c"
                /db_xref="PID:e1296647"
                /db_xref="PID:g3192003"
                /translation="MSEIPAIRRRRASQFDHIDCGRADECPVHEPLLGHPTRKVPSQL
                DGAMTLAILAVAGVVSIFLFALKGVLDQVPDVIDSASRARDAWQRFKRKANEPPSPTV
                LRELPPPAEDEQEPPAAA"
gene            23631. .23846
                /gene="SC2E1.28"
CDS             23631. .23846
                /gene="SC2E1.28"
                /note="SC2E1.28, possible prophage transcriptional
                regulator, len: 71 aa; similar to the N-terminus of many
                e.g. TR:Q37946 Salmonella typhimurium bacteriophage L
                repressor protein C2 (220 aa), fasta scores; opt: 127
                z-score: 227.1 E():2e-05, 45.7% identity in 46 aa overlap.
                Contains probable helix-turn-helix motif at aa 21 to 42
                (Score 1829, +5.42 SD)"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /db_xref="PID:e1296648"
                /db_xref="PID:g3192004"
                /translation="MPARTFSPDALRQIRRQRGVTQYKLAASLGVHNTAISHYENGHR
                RPDVETLATIADVLGARMDDFVTARAA"
RBS             23996. .24001
                /note="possible RBS upstream of SC2E1.29"
gene            24008. .26278
                /gene="SC2E1.29"
CDS             24008. .26278
                /gene="SC2E1.29"
                /note="SC2E1.29, unknown prophage ATP binding protein,
                len: 756 aa; some similarity to TR:Q38030 (EMBL:X76288)
                Streptomyces Bacteriophage phi-C31 early region ORF 9 (519
                aa), fasta scores; opt: 360 z-score: 374.4 E(): 1.3e-13,
                24.4% identity in 553 aa overlap, and to a putative
                prophage ORF from M. tuberculosis TR:O06608 (EMBL:Z95586)
                MTCY336.22 (471 aa), fasta scores; opt: 655 z-score: 535.5
                E(): 1.4e-22, 29.8% identity in 416 aa overlap. Contains
                2x PS00017 ATP/GTP-binding site motif A (P-loop)"
                /codon_start=1
                /transl_table=11
                /product="putative ATP binding protein"
                /db_xref="PID:e1296649"
                /db_xref="PID:g3192005"
                /translation="MSYTTSEARDLAATQSGLDVPEHDTDRAAYFARLGFGVFPMPTN
                KKMPRAKGWPDLAMTTVEEVEEMWRSFTPSGWQNIALTLDGWTVVDVDPKNGGTASLE
                KLTAEYELPVTRTHHTASGPDSLHFIYRADPDRPIKSGPLSSRYPGIDIKTGRGSLIV
                APGSVINGRRYEVVSAAEPAVAPSWLSEIRGDAQYPVVPSARSRTPSRASAGHSRAAT
                LAELRALPPDDPGRGNGWLTQVAGHLARNHTDHDSYLHELRAIDSESEVPHDKDRFMK
                TADSIWEREQAKERPKAAAPDKPKSILWHHEHTDLGNGRRLRDLHGEDMRHAPGWGWL
                TWDGRRWARGDGQARRFMHSVADAISEAIESMEEKQRSDDLMRWFGRSCSGPGISAAL
                NEASVMLPIETEISDFDADPHKLLVGNGVVDLRTGELLPVDRKYLLTRGTTVEYDPNA
                DCPMWMDFLGWAFQGDIEMIEYIQRMFGMCLIGNNAHQVAFFLYGPGRSGKTTLTRVL
                SRLLGDYATSADLSVFNESSSGHNEPLARLAGARLVVFSETRQGQRINEAQFKKFTGE
                DTLTASYKNKSWFEFLPAFTPVMFGNAQPSIAFDSGVERRMKVIPMRAQISDGQKNPK
                LVEQMMLNEGPAIMAWAVEGARLTAAESFVPDPPQVAQAVKEYKRENDHIGDFIEECL
                VFDEGATVASDVVWGRYSKWIEAGGVDFVKKADDRAGKSLLVRMLKTWSIENGTPIES
                FKSSNTRKLRGVALND"
misc_feature    25490. .25513
                /gene="SC2E1.29"
```

misc_feature 26147. .26170
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
/gene="SC2E1.29"
RBS 26551. .26554
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
gene 26562. .27113
/note="possible RBS upstream of SC2E1.30"
CDS 26562. .27113
/gene="SC2E1.30"
/gene="SC2E1.30"
/note="SC2E1.30, unknown prophage gene, len: 183 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.30"
/db_xref="PID:e1296650"
/db_xref="PID:g3192006"
/translation="MNDLDRRGGRIFGRVDGEVDWPSGVDSPYRLMKLRDPEKYEREV
ARQVDAQRVIVEWAEWYGLKRAKMACCPRWLTHKVSRRCNSRCTAFGGGATDHRWLEH
GVGWVLDGQPVALTSASYGISDEYRARFEWWVNEDPRLAVALGGQGWYGYGSTQVLVW
RTDVIEVMEPAPESVRALGYPQY"
gene 27578. .27769
/gene="SC2E1.31"
CDS 27578. .27769
/gene="SC2E1.31"
/note="SC2E1.31, possible prophage transcriptional
regulator, len: 63 aa; similar to the N-terminus of many
e.g. TR:E264367 (EMBL:X87420) Salmonella typhimurium
bacteriophage ES18 gene C2 (216 aa), fasta scores; opt:
133 z-score: 207.9E(): 0.00024, 35.8% identity in 53 aa
overlap. Contains probable helix-turn-helix motif at aa 17
to 38 (Score 1786, +5.27 SD)"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="PID:e1296651"
/db_xref="PID:g3192007"
/translation="MFSGRKLKEARNKRGMSQAAVAARMGLSGLYVSYWENGRYAPNG
VNMVLLLRALGCELEEVTE"
RBS 27754. .27759
/gene="SC2E1.31"
gene 27766. .28116
/note="possible RBS upstream of SC2E1.32"
CDS 27766. .28116
/gene="SC2E1.32"
/gene="SC2E1.32"
/note="SC2E1.32, unknown prophage gene, len: 116 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.32"
/db_xref="PID:e1296652"
/db_xref="PID:g3192008"
/translation="MNAPETEFADAGWVAITRDEGGRVGVKPVRMFELTGEGVTALTK
NADGLMVPDPHAVEIINTDPLHAAKLAWLHKLVGVAERCTTDEARADLRRIAEWLIDW
EPGDPGLRLADAEA"
gene 28184. .28672
/gene="SC2E1.33"
CDS 28184. .28672
/gene="SC2E1.33"
/note="SC2E1.33, unknown prophage gene, len: 162 aa;
contains probable helix-turn-helix motif at aa 11 to 32
(Score 1192, +3.25 SD)"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.33"
/db_xref="PID:e1296653"
/db_xref="PID:g3192009"
/translation="MVVVLAQQGIGRNEISRRTGISTASVSRIAAAHGIEFDGSQTEV
ALKARIAELKARQAGIALGLVEDITSARGLLRMATTHRDFAFAAKAISDLTQSAQRMT
PEITNEEELEEAKEGLRDLYFQMGLMREGFEAKYGVPLDSDEGRELWDKARQEDMENE
QP"
RBS 28644. .28649
/gene="SC2E1.33"
/note="possible RBS upstream of SC2E1.34"
gene 28659. .29075
/gene="SC2E1.34"
CDS 28659. .29075
/gene="SC2E1.34"
/note="SC2E1.34, unknown prophage gene, len: 138 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.34"
/db_xref="PID:e1296654"
/db_xref="PID:g3192010"
/translation="MSNRDFSEDVQGPRSTDGPSVDDLRAEMYGGQVAGRFHYRESRT
ERDSAVIQGTPVLGDGTMTDDQAAEIRKSFGIGADLRHAAQAAESYMDRRDPTSGHTI
RTYGNGRVVDLDAEPTQTGSDRVEYPRVWYGDDPGA"
RBS 29481. .29484
/note="possible RBS upstream of SC2E1.35"
gene 29493. .29936
/gene="SC2E1.35"
CDS 29493. .29936
/gene="SC2E1.35"
/note="SC2E1.35, unknown prophage gene, len: 147 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.35"
/db_xref="PID:e1296655"
/db_xref="PID:g3192011"
/translation="MAATKVFRPEFKTRDGRKVSFGANYTHDTKAKRRAGLTTAGRPK
VETVGAGTVVRSPRRWLPWAVFGTNSAVLGGTWAVAELGPAVGGLVVGVALWAGYRIA
RWTAGEWAYFATKSRQIRGGRQVAEFASEGLPPQAHIPGEKVTRP"
RBS 30237. .30240
/note="possible RBS upstream of SC2E1.36"
gene 30248. .30754
/gene="SC2E1.36"
CDS 30248. .30754
/gene="SC2E1.36"
/note="SC2E1.36, unknown prophage gene, len: 168 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.36"
/db_xref="PID:e1296656"
/db_xref="PID:g3192012"
/translation="MAAKKTETAEATPCECSMYDALPADLTEEQVASGDFEVLTTGCT
ATTKRQFAPGHDAKLKSALIRWGALGLEIRRNEGGVATSASPAKHAARYAFAHMVTAG
VKRAEAKAAEKAERAAARAAKKAAPAPEVIKAKVGRVTYQGRMDGDHFVYEVKGQERR
TLKFQPAA"
gene 30824. .33214
/gene="SC2E1.37"
CDS 30824. .33214
/gene="SC2E1.37"
/note="SC2E1.37, unknown prophage gene, len: 796 aa; shows
very limited similarity to Rhodococcus opacus TR:G2935035
(EMBL:AF003948) putative transposase homolog (242 aa),
fasta scores; opt: 215 z-score: 280.3 E(): 2.2e-08, 33.3%
identity in 105 aa overlap. Contains PS00397 Site-specific
recombinases active site and Pfam match to entry
recombinase PF00239, Site-specific recombinases, score
36.16. The last part of this ORF (aa 672 onwards) appears
to be a continuation of SC2E1.21, which was interrupted by
the prophage; the similarities to YX27_MYCTU MTCY274.27c
(389 aa) (fasta scores; opt: 246 z-score: 225.9 E():
2.4e-05, 38.8% identity in 116 aa overlap) continue after
aa 672. Contains probable coiled-coil from aa 449 to 505"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.37"
/db_xref="PID:e1296657"
/db_xref="PID:g3192013"
/translation="MYQRNTDALDVAMGVVERAVQPVRLRAVAYARVSTEEQAKGYGV
QAALKKILRYIGRKDWDHVGTYTDEGISGSLEAADREDLKRLMADAHRQPKPFDIVVV
SEGRAIGRVGRAFWRWVWALEDIGIYVGVVEDDYDNSTPEGRKKMRRDADYAETEWET
IRKRTQGGLQEKAEDGGWPGGQPPYGYEIACQGQKGKSHLVQCAAEVKTLRMAWAMVV
EEKLNTREVAARLNALNRFTRSGVPWSHSNLRAKLLSESTINARVIFRNPNRAHAGHG
AKFGEDGKPLHGETVIIKLKPIFEPGEVAALQVTLAQTSKGAGRPKPKPYPLSKRLFN
EHTGCNSHHVGMARTSREGRWYRCTGLAAKYPGDPTCTCKMVDADAMESAVWGEVVSL

LGDPDRLRAMAAEWVGMAAGDQVQHADRIADFDKQIANLDRAIASMATECVKAGLPAA
AIAQASAALLEERRQLADLRDEAAAWLEETEAAEQRARDLEALATAARTRLADMDPAE
QGEVLALLDVKVTITGPVPKPKLGLTCSLAEWFKVAGRLVPSELTDDMWAAAEPVVKA
WEPSNHKLRPGRLMLDAMFYKARTGCRWDDLPERFGPWKGIHSRYKTWRNCGVWDEIM
AALPVDGPGYRPVPEINLVPPFRVEGRVDPRVLAGPDIQEEVVRPETGVPGPATSALS
AGVHEMLRGDAVLVGDAAEVVELVGAMGELPPERRGPVLPTDLLESRTRQVLAGLPGR
GTATAGEVALRAQTTQDDAVARLYELRALGYVERHGDGWKLTRQAMISVRGGRSPC"

misc_feature 30905. .31387
/gene="SC2E1.37"
/note="Pfam match to entry recombinase PF00239, Site-specific recombinases, score 36.16"

misc_feature 30911. .30937
/gene="SC2E1.37"
/note="PS00397 Site-specific recombinases active site"

gene 33497. .34339
/gene="whiG"

CDS 33497. .34339
/gene="whiG"
/note="SC2E1.38, whiG, RNA polymerase sigma factor WhiG, len: 280 aa; identical to RPSW_STRCO RNA polymerase sigma factor whiG (280 aa). Also highly similar toRPSD_BACSU RNA polymerase sigma-D factor (sigma-28) (254 aa)(E(): 0, 43.0% identity in 242 aa overlap) and FLIA_ECOLI RNA polymerase sigma factor for flagellar operon (sigma F) (239 aa)(E(): 0, 42.8% identity in 229 aa overlap). Contains PS00715 and PS00716 Sigma-70 factors family signatures 1 and 2, Pfam match to entry sigma70 PF00140, Sigma-70 factors, score 342.03 and probable helix-turn-helix motif at aa 245 to 266 (Score 1657, +4.83 SD)"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor WhiG"
/db_xref="PID:e1296658"
/db_xref="PID:g3192014"
/translation="MPQHTSGSDRAAIPPAARDGGSVRPPAPSTLDELWRSYKTTGDE
RLREQLILHYSPLVKYVAGRVSVGLPPNVEQADFVSSGVFGLIDAIEKFDVDREIKFE
TYAITRIRGAMIDELRALDWIPRSVRQKARNVERAYATLEARLRRTPSESEVAVEMGI
AVEDLHAVFSQLSLANVVALEELLHAGGEGGGRLSLMDTLEDTAADNPVEVAEDRELR
RLLARAINTLPEREKTVVTLYYYEGLTLAEIGNVLGVTESRVSQIHTKSVLQLRAKLA
GFGR"

misc_feature 33626. .34315
/gene="whiG"
/note="Pfam match to entry sigma70 PF00140, Sigma-70 factors, score 342.03"

misc_feature 33725. .33766
/gene="whiG"
/note="PS00715 Sigma-70 factors family signature 1"

misc_feature 34229. .34309
/gene="whiG"
/note="PS00716 Sigma-70 factors family signature 2"

RBS 34449. .34453
/note="possible RBS upstream of SC2E1.39"

gene 34468. .35025
/gene="SC2E1.39"

CDS 34468. .35025
/gene="SC2E1.39"
/note="SC2E1.39, possible transcriptional regulator, len: 185 aa; equivalent to Streptomyces roseosporus TR:P72569 (EMBL:U58281) tetR homologue whiR (179 aa) (86.6% identity in 179 aa overlap) and similar to several putative transcriptional regulators of the tetR family e.g. TR:E1247678 putative transcriptional regulator SC10A5.11 (S.coelicolor) (198 aa), fasta scores; opt: 434 z-score: 720.5 E(): 6.8e-33, 41.5% identity in 171 aa overlap. Contains probable helix-turn-helix motif at aa 29 to 50 (Score 1618, +4.70 SD) and Pfam match to entry tetR PF00440, Bacterial regulatory proteins, tetR family, score 29.18"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="PID:e1296659"
/db_xref="PID:g3192015"
/translation="MAEHRSMQRAALLDAARSLLSDGGTEALTFPALAERTGLARSSV
YEYFRSRAAVVEELCAVDFPVWAAEVEAAMERETAPDAKVEAYVRTQLDLVGDRRHRA
VVAISASELDAGAREKIRAAHGGLIAMIVEALGEMGHGQPRLAAMLLQGVVDAAVRRI
ELGAAEEPGAITEAAVSMALRGVRG"

misc_feature 34501. .34641
/gene="SC2E1.39"
/note="Pfam match to entry tetR PF00440, Bacterial regulatory proteins, tetR family, score 29.18"

gene complement(34979. .35659)
/gene="SC2E1.40c"

CDS complement(34979. .35659)
/gene="SC2E1.40c"
/note="SC2E1.40c, probable secreted protein, len: 226 aa; some similarity to two putative secreted M. tuberculosis proteins YX22_MYCTU MTCY274.22 (249 aa), fasta scores; opt: 315 z-score: 268.8 E(): 9.7e-08, 40.7% identity in 145 aa overlap (also upstream of rpsB) and Y0GK_MYCTU MTCY13D12.20 (407 aa), fasta scores; opt: 218 z-score: 176.1 E(): 0.014, 34.1% identity in 164 aa overlap. Contains probable N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.40c"
/db_xref="PID:e1296660"
/db_xref="PID:g3192016"
/translation="MRARARVVVLSAVLAVLAVPTAQAGRAAGPPPPATAAAAGWALP
YAVGPMGRVDPAVPPVARHWPVGTRPAVLRGWDPPANPYGPGHRGVDLAAGGGAPVRA
VAAGRVSFAGPVGGRGVVSVELTGTGEPPLRTTFAPVDAAVRKGDEVAAGEPLGTVQP
TGSHCTACLHWGLLRGEVYLNPLSLLPPWLLDAGPSRLLPVAGVPVPGRGGISPVRRA
TPWRPPPP"

misc_feature complement(35024. .38962)
/note="overlap with cosmid 6A9 (unfinished)"

RBS complement(35666. .35669)
/note="possible RBS upstream of SC2E1.40c"

misc_feature 35757. .37919
/note="97.4% match to AF034101 AF034101 Streptomyces coelicolor ribosomal protein S2 (rpsB) and elongation factor Ts (tsf) genes from 1 to 2154"

RBS 35934. .35939
/note="possible RBS upstream of rpsB"

gene 35948. .36880
/gene="rpsB"

CDS 35948. .36880
/gene="rpsB"
/note="SC2E1.41, rpsB, 30S ribosomal protein S2, len: 310 aa; almost identical to RS2_STRCO 30S ribosomal protein S2 (242 aa), E(): 0, 96.7% identity in 242 aa overlap and also highly similar to many others e.g. RS2_ECOLI 30S ribosomal protein S2 (240 aa), fasta scores; opt: 744 z-score: 950.7 E(): 0, 50.4% identity in 224 aa overlap. Contains PS00962 Ribosomal protein S2 signature 1 and PS00211 ABC transporters family signature (probably a false positive). Also contains Pfam matches to entry S2 PF00318, Ribosomal protein S2, score 304.89 and to entry 60s_ribosomal PF00428, 60s Acidic ribosomal protein, score 24.36. Note that this sequence is longer than RS2_STRCO due to a sequence difference around 235 aa. This sequence has a C-terminal extension not present in other rpsB. The M. tuberculosis protein also has a C-terminal extension but with a differentsequence. Our sequence has been checked, and is also confirmed by sequence from the overlapping cosmid 6A9."
/codon_start=1
/transl_table=11
/product="30S ribosomal protein S2"
/db_xref="PID:e1296661"
/db_xref="PID:g3192017"
/translation="MAVVTMRELLESGVHFGHQTRRWNPKMKRFIFTERNGIYIIDLL
QSLSYIDRAYEFVKETVAHGGTVMFVGTKKQAQEAIAEQATRVGMPYVNQRWLGGMLT
NFSTVYKRLQRLKELEQIDFEDVAASGLTKKELLVLSREKAKLEKTLGGIREMSKVPS
AVWIVDTKKEHIAVGEARKLNIPVVAILDTNCDPDEVDYKIPGNDDAIRSVTLLTRVI
ADAVAEGLIARSGAAGGAKGDKAAGEPLAAWERDLLEGEKAEKKDDAEAAEKPAEAPA

```
                AEAPAAEAAEAPAAEAAPAEEPAAEAPAADAEQA"
misc_feature    35963. .35998
                /gene="rpsB"
                /note="PS00962 Ribosomal protein S2 signature 1"
misc_feature    35972. .36625
                /gene="rpsB"
                /note="Pfam match to entry S2 PF00318, Ribosomal protein
                S2, score 304.89"
misc_feature    36461. .36505
                /gene="rpsB"
misc_feature    36740. .36856
                /gene="rpsB"
                /note="PS00211 ABC transporters family signature"
gene            36975. .37811
                /gene="tsf"
CDS             36975. .37811
                /gene="tsf"
                /note="SC2E1.42, tsf, elongation factor Ts, len: 278.
                Almost identical to EFTS_STRCO elongation factor TS
                (EF-TS) (276 aa), fasta scores; opt: 1695 z-score: 2425.0
                E(): 0, 97.8% identity in 278 aa overlap and highly
                similar to many others e.g. EFTS_ECOLI (282 aa), fasta
                scores; opt: 321 z-score: 681.2 E(): 1e-30, 38.4% identity
                in 276 aa overlap. Contains PS01126 Elongation factor Ts
                signature 1"
                /codon_start=1
                /transl_table=11
                /product="elongation factor Ts"
                /db_xref="PID:e1296662"
                /db_xref="PID:g3192018"
                /translation="MANYTAADVKKLRELTGAGMMDCKKALDEAEGNVEKAVEALRIK
                GQKGVAKREGRSAENGAVVSIIADDNTSGVLVELKCETDFVAKGEKFQNVATAIAEHV
                AKAAPADLEALLASEIEPGKTVQAYVDEANANLGEKIVLDRFAQFSGGFVTAYMHRTM
                PDLPPQIGVLVELDKPNAEIAKGVAQHIAAFAPKYLSKEDVPAEVVESERRVAEETTR
                AEGKPEAALPKIVEGRLNGFFKDATLLGQPYALDNKKSVQKVLEEAGVSLKRFSRIKV
                GI"
misc_feature    37008. .37055
                /gene="tsf"
                /note="PS01126 Elongation factor Ts signature 1"
RBS             38004. .38007
                /note="possible RBS upstream of pyrH"
gene            38017. .38778
                /gene="pyrH"
CDS             38017. .38778
                /gene="pyrH"
                /note="SC2E1.43, pyrH, uridylate kinase, len: 253 aa;
                highly similar to many e.g. PYRH_ECOLI uridylate kinase
                (EC 2.7.4.-) (240 aa), fasta scores; opt: 668 z-score:
                968.6 E(): 0, 45.9% identity in 233 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="uridylate kinase"
                /db_xref="PID:e1296663"
                /db_xref="PID:g3192019"
                /translation="MSSTKAEKSDDGKVSGRFMLKLSGEAFSGGGGLGVDPDVVHAIA
                REIAAVVRDGAQIAIVIGGGNFFRGAELQVRGMDRARSDYMGMLGTVMNCLALQDFLE
                KEGVDCRVQTAITMGQVAEPYIPLRAVRHLEKGRVVIFGAGMGMPYFSTDTTAAQRAL
                EIDAEALLMGKNGVDGVYDSDPKTNPDAVKFDALGYGEVITRDLKVADATAVTLCRDN
                SLPIVVFELLKEGNIARAVKGEKIGTLVGDQGSRD"
BASE COUNT      6233 a  13696 c  13758 g   5275 t
ORIGIN
```

```
  Query Match              7.0%;  Score 38.2;  DB 1;  Length 38962;
  Best Local Similarity   54.7%;  Pred. No. 0.52;
  Matches   76;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy      46 GTGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAG 105
           | || ||| ||    | |  |  ||| || | | | |||   |    || | |    ||
Db   35657 GAGGGCGCGTGCGCGAGTGGTGGTGCTGTCGGCCGTCCTGGCCGTCCTGGCGGTCCCGAC 35598

Qy     106 TGCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTG 165
           ||  | || | ||| || ||  |     | |  | || |||| | ||| ||||| | |||
Db   35597 CGCGCAGGCCGGCCGGGCCGCGGGCCCGCCGCCCCCGGCGACGGCGGCGGCTGCGGGGTG 35538

Qy     [illegible]
           || |||| |  || ||| |
Db   35537 GGCCCTGCCCTACGCGGTG 35519

RESULT   8
SC2E1/c
LOCUS       SC2E1       38962 bp    DNA                 BCT       04-JUN-1998
DEFINITION  Streptomyces coelicolor cosmid 2E1.
ACCESSION   AL023797
NID         g3191976
KEYWORDS    30S ribosomal protein S16; 30S ribosomal protein S2; 50S ribosomal
            protein L19; cell division protein FtsH homolog; elongation factor
            Ts; ffh; ftsH; glnB; glnD; nitrogen regulatory protein pII;
            prophage; protein pII uridylyltransferase; pyrH; RNA polymerase
            sigma factor WhiG; rplS; rpsB; rpsP; signal peptidase I; signal
            recognition particle protein; sipl; sip2; sip3; sip4; trmD; tRNA
            (guanine-N1)-methyltransferase; tsf; uridylate kinase; whiG.
SOURCE      Streptomyces coelicolor.
  ORGANISM  Streptomyces coelicolor
            Eubacteria; Firmicutes; Actinomycetes; Streptomycetes;
            Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 38962)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 38962)
  AUTHORS   Murphy,L. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 38962)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
            correct initiation codon. Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon). If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
```

overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2E1 lies between 7A1 and 6A9 on the AseI-B genomic restriction fragment.

FEATURES Location/Qualifiers

source 1. .38962
/organism="Streptomyces coelicolor"
/strain="A3(2)"
/db_xref="taxon:1902"
/clone="cosmid 2E1"

gene 1. .166
/gene="glnB"

CDS <1. .166
/gene="glnB"
/note="SC2E1.01, glnB, nitrogen regulatory protein pII, partial CDS, len: >54 aa; highly similar to many e.g. GLNB_AZOBR (112 aa), fasta scores; opt: 198 z-score: 381.9 E(): 4.9e-14, 50.0% identity in 54 aa overlap"
/codon_start=2
/transl_table=11
/product="nitrogen regulatory protein pII"
/db_xref="PID:e1296621"
/db_xref="PID:g3191977"
/translation="IRIEVLVEDDDAEQLIDVVVKAARTGKIGDGKVWAVPVDTAVRV RTGERGPDAL"

gene 231. .2738
/gene="glnD"

CDS 231. .2738
/gene="glnD"
/note="SC2E1.02, probable glnD, protein pII uridylyltransferase, len: 835 aa; similar to many e.g. GLND_ECOLI (890 aa), fasta scores; opt: 403 z-score: 540.4 E(): 7.3e-23, 26.2% identity in 851 aa overlap. Contains 7 degenerate repeats of the sequence PSS"
/codon_start=1
/transl_table=11
/product="putative protein pII uridylyltransferase"
/db_xref="PID:e1296622"
/db_xref="PID:g3191978"
/translation="MTDEAEDSGPGGYAAARLRLLTEGARSGPPRRRALAELTDGWLA GLFGAATEEHTGISLVAVGGYGRGELSPRSDLDLLLLHDGRDDKAVAALADRLWYPVW DLGIDLDHSVRTPQQARKTAGQDLKVHLGLLDARHLAGDLGLTASLRTAVLADWRNQA PKRLPELRDLCAERAERQGELQFLLEPDLKEARGGLRDATALRAVAASWLADAPREGL AEARRRLLDVRDALHLATGRATDRLALQEQDQVAAELGLLDADALLRQVYEAARVISY AGDVTWREVGRVLRSRSVRPRLRAMMNGRNGGKPVAERSPLAEGVVEQDGEAVLARTA RPERDPALPLRAAAAAAQAGLPLSRHAVRRLAATARPLPTPWPAEAREQLVTLLGSGR PTVQVWEALEAEGLVTRLLPDWERVRCRPQRNAVHLWTVDRHLIETAVRAAGFTRRVH RPDLLLIAALLHDIGKGWPGDHSVAGETIARDVAARIGFDGADTAVLATLVRHHLLLV ETATRRDLDDPATVRAVAQAVGTEHTLELLHALTEADALATGPAAWSSWRGSLVADLV KRVSGVLAGEPQPEAESAAPTAEQERLAVEAFRTGGPVLALRAQTEPPADSAPAPSSP SSPSFPSPLSSPSSPSSADGPEPLGVELLIAVPDQAGVLPAVAGVLAMHRLTVRTAEL RSVPLPDGVEGSVLLLDWRVAAQYGSLPQAARLRADLVRALDGTLDIAARLAERDAAH PRRRGVEPPPPRVTVAPAASRLATVIEVRAQDAPGLLFRLGRALEAAGVRVRSAHVST LGANAVDAFYVTRGEGTPLPGDEAASVARGLEESLRT"

repeat_region 2091. .2153
/note="7 degenerate 9bp repeats"

RBS 2719. .2723
/gene="glnD"
/note="possible RBS upstream of ffh"

gene 2735. .4387
/gene="ffh"

CDS 2735. .4387
/gene="ffh"
/note="SC2E1.03, ffh, signal recognition particle protein, len: 550 aa; highly similar to mnay e.g. SR54_ECOLI (453 aa), fatsa scores; opt: 1465 z-score: 1282.8 E(): 0, 50.0% identity in 458 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00300 SRP54-type proteins GTP-binding domain signature and Pfam match to entry SRP54 PF00448, SRP54-type proteins, score 66.57. Similarities suggest possible start site at aa 35"
/codon_start=1
/transl_table=11
/product="signal recognition particle protein"
/db_xref="PID:e1296623"
/db_xref="PID:g3191979"
/translation="MTVRSVQRAGTLCLPGRIPWRTTHTNADARTYAAVFDTLSDRLS ATFKSLRGKGRLTEADIDATAREIRIALLEADVALPVVRAFIKKVKERSLGAEVSKAL NPAQQVLKIVNEELVGILGGETRRLRFAKQPPTVIMLAGLQGAGKTTLAGKLGHWLKE QGHSPLLVACDLQRPNAVNQLSVVAERAGVAVYAPEPGNGVGDPVKVAKDSIEFAKAK VHDLVIVDTAGRLGIDQELMQQAADIRDAVSPDEILFVVDAMIGQDAVNTAEAFRDGV GFDGVVLSKLDGDARGGAALSIASVTGKPIMFASNGEKLDDFDAFHPDRMASRILDMG DLLTLIEQAEKTFSQEEAEKMASKLASKKGQDFTLDDFLAQMEQVRKMGSISKLLGML PGMGQMKDQINNLDERDVDRTAAIIKSMTPGERQEPTIINGSRRARIAKGSGVEVSAV KNLVERFFEARKMMSRMAQGGGMPGMPGMPGMGGGPGRQKKQQKKAKGKQRSGNPMKR KQQEQEEEAARRAAAAQSGGALGLPQQGGQDFELPDEFKKFMG"

misc_feature 3155. .3178
/gene="ffh"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

misc_feature 3566. .3697
/gene="ffh"
/note="Pfam match to entry SRP54 PF00448, SRP54-type proteins, score 66.57"

misc_feature 3647. .3688
/gene="ffh"
/note="PS00300 SRP54-type proteins GTP-binding domain signature"

stem_loop 4405. .4453
/note="hairpin loop with 22/23 bp stem"

gene 4504. .6450
/gene="ftsH"

CDS 4504. .6450
/gene="ftsH"
/note="SC2E1.04, ftsH, cell division protein FtsH homolog (zinc metalloprotease, integral membrane protein), len: 648 aa; similar to many e.g. FTSH_LACLA (695 aa), fasta scores; opt: 1489 z-score: 1195.5 E(): 0, 41.7% identity in 640 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00674 AAA-protein family signature and Pfam matches to entry AAA PF00004, ATPases associated with various cellular activities (AAA) and to entry zn-protease PF00099, zinc-binding metalloprotease domain, score 22.07, score 357.96. Proline-rich N-terminus is not present in other ftsH homologs"
/codon_start=1
/transl_table=11
/product="cell division protein FtsH homolog"
/db_xref="PID:e1296624"
/db_xref="PID:g3191980"
/translation="MTNPSPPRKAPEPPWRTEGTPDEPPKPPPGGRRMRGGWWNLILA ALIVYLIANLVLSFFNEGDEPTISYTEFSKQVDEGNVSKIYAKGDAIQGQLKKARDNP EGDGTYTKFTTERPTFADDQLWADLTKNRVTVTAEPVVQHRSFLANLLIALAPMLILV VLWIFIARRMRGALGGGAGGMLGRKAPPKPVELEAGKPRTTFADVAGIDEVEGELSDV VDFLKNPDAYRRMGAKMPRGVLLTGPPGTGKTLLARAVAGEAGVPFFSASASEFIEMI VGVGASRVRELFAEARKVAPSIIFIDEIDTIGRARGGGSGTGGHDEREQTLNQILTEM DGFSGSEGVIVIAATNRADILDAALTRPGRFDRVVSVSPPDRGGREAILEIHTREIPL APDIDLAQVARTTPGMTGAELANLANEAALLAVKRKQERVTQANLSEALEKVQLGAER PLVMPEEERRRTAYHESGHALLGMLQPGADPVRKITIVPRGRALGVTLSTPDADKYAY TEEYLRGRIIGALGGMAAEHVVYGMITTGSESDLEQVTNIARGMVARWGMSERVGRLS ALPGDAQQAYGLAAAPQTLDAIDGEMRRVVDSCYEEAVRKLRDHRGQLDALAESLLAS ETLDEADAYRIAGITRLTKDDPEA"

misc_feature 5215. .5781
/gene="ftsH"
/note="Pfam match to entry AAA PF00004, ATPases associated with various cellular activities (AAA), score 357.96"

misc_feature 5230. .5253
/gene="ftsH"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

misc_feature 5530. .5586
/gene="ftsH"
/note="PS00674 AAA-protein family signature"

misc_feature 5884. .5931
/gene="ftsH"
/note="Pfam match to entry zn-protease PF00099, zinc-binding metalloprotease domain, score 22.07"

gene complement(6468. .7085)
/gene="SC2E1.05c"

```
CDS             complement(6468. .7085)
                /gene="SC2E1.05c"
                /note="SC2E1.05c, probable lipoprotein, len: 205 aa;
                contains N-terminal signal sequence and appropriately
                positioned PS00013 Prokaryotic membrane lipoprotein lipid
                attachment site"
                /codon_start=1
                /transl_table=11
                /product="putative lipoprotein"
                /db_xref="PID:e1296625"
                /db_xref="PID:g3191981"
                /translation="MPHRTTALRLAAPVAVLALALTACGGGDDDSGTDAKGGATTEAT
                TQASPSPSQPASGGDMAQGESASGKVAEGPGEVTYDVLAQKVEVGTEAEAAKAVADAA
                DAKGKVLATAHIKYTHQGGPALTEGSDVHDGTTVFADGQRGGVLIGAAEDAPGCEDPY
                DAESWKQGESHVFCETYVIPAGATSVEVHWSEEDGEPYVWKFPQK"
misc_feature    complement(7014. .7046)
                /gene="SC2E1.05c"
                /note="PS00013 Prokaryotic membrane lipoprotein lipid
                attachment site"
RBS             complement(7092. .7096)
gene            /note="possible RBS upstream of Sc2E1.05c"
                complement(7236. .8123)
CDS             /gene="SC2E1.06c"
                complement(7236. .8123)
                /gene="SC2E1.06c"
                /note="SC2E1.06c, unknown, len: 295 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.06c"
                /db_xref="PID:e1296626"
                /db_xref="PID:g3191982"
                /translation="MTPTLVRQHSSHAQATPHVDPRARARDWSEIQERMLVPLYEAVY
                ERLDVGGDTRMLGLNCGSGLALLMAASRGAAVTGVDSRSPELLDLARRRLLSGTSGDP
                GADSTVSTRTRGDAPRIVQGTPGDAADPTAPAFTLVTAFEPGGCRAGDSETLGDLLAR
                ATPLAARGAAVVLAGWGPPERCATTSVLRVAAKLTDPTHGTGGWRPALRDDLEEVAQR
                AGLRPDGSGRVACPFGYADTGSAVRGLLSTGMFDAAVAATDRDQVDKELTEALHPYQR
                PDGTVWMPNVFRYLIARVR"
gene            8793. .9389
                /gene="SC2E1.07"
CDS             8793. .9389
                /gene="SC2E1.07"
                /note="SC2E1.07, unknown, len: 198 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.07"
                /db_xref="PID:e1296627"
                /db_xref="PID:g3191983"
                /translation="MSTTPPVLAAELAQAWADIQRHHTELPDLAAPESLIGESSSACG
                HELSFERLLHEAVHGIAAARGVRDTSRAGRYHNRRFLAIAEELGLDHPEEPHPSSGFS
                LVTLTPEAKRRYRPTIERLQRALKAHTVATATDTARSFRGPAARHGSSGGGVRVKAVC
                DCGRNVRVVPSVLAQAPIMCGGCGKPFRIPEVVGAGMS"
RBS             9629. .9633
gene            /note="possible RBS upstream of rpsP"
                9645. .10064
CDS             /gene="rpsP"
                9645. .10064
                /gene="rpsP"
                /note="SC2E1.08, rpsP, 30S ribosomal protein S16, highly
                similar to many e.g. RS16_MYCTU (162 aa), fatsa scores;
                opt:480 z-score: 771.3 E(): 0, 56.3% identity in 144 aa
                overlap. Contains PS00732 Ribosomal protein S16 signature"
                /codon_start=1
                /transl_table=11
                /product="30S ribosomal protein S16"
                /db_xref="PID:e1296628"
                /db_xref="PID:g3191984"
                /translation="MAVKIKLKRLGKIRSPHYRIVVADSRTRRDGRAIEEIGKYHPTY
                NPSVMEVDAERVAYWLGVGAQPTEPVLAILKKTGDWQKFKGEPAPAPLLQPAEKAARP
                SFEAIGGEDEGKGEAITQKKKADKKDEAAAESSASEA"
misc_feature    9651. .9680
                /gene="rpsP"
                /note="PS00732 Ribosomal protein S16 signature"
RBS             10056. .10059
                /gene="rpsP"
gene            /note="possible RBS upstream of SC2E1.09"
                10067. .10306
CDS             /gene="SC2E1.09"
                10067. .10306
                /gene="SC2E1.09"
                /note="SC2E1.09, unknown, len: 79 aa; highly similar to
                small hypothetical proteins from several organisms, e.g.
                YX40_MYCTU MTCY274.40c (80 aa), fasta scores; opt: 356
                z-score:628.4 E(): 9.1e-28, 71.6% identity in 74 aa
                overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.09"
                /db_xref="PID:e1296629"
                /db_xref="PID:g3191985"
                /translation="MLEEALEHLVKGIVDNPDDVQVASRNLRRGRVLEVRVHPDDLGK
                VIGRNGRTARALRTVVGAIGGRGVRVDLVDVDHVR"
RBS             10403. .10407
gene            /note="possible RBS upstream of SC2E1.10"
                10418. .10984
CDS             /gene="SC2E1.10"
                10418. .10984
                /gene="SC2E1.10"
                /note="SC2E1.10, unknown, len: 188 aa; similar to
                hypothetical proteins from many organisms e.g. YX38_MYCTU
                MTCY274.38c (176 aa), fasta scores; opt: 479 z-score:
                693.1 E(): 2.3e-31, 48.3% identity in 176 aa overlap. Also
                some similarity to RIMM_HAEIN 16S rRNA processing protein
                rimM (178 aa), fasta scores; opt: 210 z-score: 174.2 E():
                0.018, 27.6% identity in 170 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.10"
                /db_xref="PID:e1296630"
                /db_xref="PID:g3191986"
                /translation="MQLVVARIGRAHGIKGEVTVEVRTDEPELRLGPGAVLATDPAST
                GPLTIESGRVHSGRLLLRFAGVHDRTGAEALRNTLLIADVDPDERPEDEDEYYDHQLI
                DLDVVTEDGTEVGRITEISHLPTQDLFVVERPDGSEVYVPFVSEIVTGIDLDAQRAVI
                DPPPGLIDDRAEIASARDAGAEDAGDEA"
RBS             10976. .10979
                /gene="SC2E1.10"
gene            /note="possible RBS upstream of trmD"
                10984. .11817
CDS             /gene="trmD"
                10984. .11817
                /gene="trmD"
                /note="SC2E1.11, trmD, probable tRNA
                (guanine-N1)-methyltransferase, len: 277 aa; similar to
                many e.g. TRMD_ECOLI (EC2.1.1.31) (255 aa), fasta scores;
                opt: 437 z-score: 660.6E(): 1.5e-29, 41.2% identity in 238
                aa overlap"
                /codon_start=1
                /transl_table=11
                /product="tRNA (guanine-N1)-methyltransferase"
                /db_xref="PID:e1296631"
                /db_xref="PID:g3191987"
                /translation="MRLDVVTIFPEYLEPLNVSLVGKARARGQLGVHVHDLRDWTYDR
                HNTVDDTPYGGGPGMVMKTEPWGDALDSVLADGYETGCGEPALVVPTPSGRPFTQELA
                VHLSERPWLIFTPARYEGIDRRVVDEYATRMPVYEVSIGDYVLAGGEAAVLVVTEAVA
                RLLPGVLGNAESHRDDSFAPGAMANLLEGPVHTKPPQWRGRGIPDVLLSGHHGKIARW
                RRDEALRRTTANRPDLIERCDPAAFDKKDREMLSILGWQPDPDGEPYGRFWRRTPGME
                E"
RBS             11939. .11942
gene            /note="possible RBS upstream of rplS"
                11954. .12304
CDS             /gene="rplS"
                11954. .12304
                /gene="rplS"
                /note="SC2E1.12, rplS, 50S ribosomal protein L19, len:
                116aa; highly similar to many e.g. RL19_MYCTU (113 aa),
                fastascores; opt: 571 z-score: 1138.4 E(): 0, 77.3%
                identity in110 aa overlap. Contains PS01015 Ribosomal
```

```
                protein L19 signature"
                /codon_start=1
                /transl_table=11
                /product="50S ribosomal protein L19"
                /db_xref="PID:e1296632"
                /db_xref="PID:g3191988"
                /translation="MSHLLDSVDAASLRSDVPAFRPGDTVNVHVRVIEGNRSRVQQFK
                GVVIRRQGAGVRETFTVRKVSFSVGVERTFPVHTPIVEKIELVTRGDVRRAKLYYLRE
                LRGKAAKIKEKRDS"
misc_feature    12215. .12256
                /gene="rplS"
                /note="PS01015 Ribosomal protein L19 signature"
gene            12350. .13129
                /gene="sip1"
CDS             12350. .13129
                /gene="sip1"
                /note="SC2E1.13, sip1, probable signal peptidase I, len:
                259 aa; similar to many e.g. LEP_HAEIN signal peptidase I
                (EC 3.4.21.89) (349 aa), fasta scores; opt: 162 z-score:
                381.2 E(): 5.3e-14, 32.5% identity in 243 aa overlap,
                though with large internal deletions. Contains PS00761
                Signal peptidases I signature 3, Pfam match to entry
                signal_pept_I PF00461, Signal peptidases I, score 51.12
                and probable transmembrane domain. Also similar to
                downstream genes sip2 (E(): 9.7e-26, 44.9% identity in 274
                aa overlap), sip3 (E(): 4.8e-16, 35.6% identity in 264 aa
                overlap) and sip4 (E(): 8.7e-08, 37.0% identity in 219 aa
                overlap)"
                /codon_start=1
                /transl_table=11
                /product="putative signal peptidase I"
                /db_xref="PID:e1296633"
                /db_xref="PID:g3191989"
                /translation="MDTETQHTERDRSSRPSDSEQPSEPEGPEERSRSAFTGRIGDWL
                PGGRITVTLLILLLFLLLVSTFVLQPFQIPSGSMERGLRIGDRVLVNKLAYRFDGRPR
                RGDIVVFDGTGSFGHGDYIKRVVGVGGDHVVCCDEEGRVRVNGQPVDESAFLYPGDRP
                STVPFDVVVPDGTLFVLGDHRSDSSDSRDHLGSPGGGMVPLDDVIGRADWIVWPFGHA
                TRLDRPDAYARVPDAEAGAGAASGAGAAPDPGAAPDGDADG"
stem_loop       12369. .12455
                /gene="sip1"
                /note="hairpin loop with 34/41 bp stem"
misc_feature    12509. .12799
                /gene="sip1"
                /note="Pfam match to entry signal_pept_I PF00461, Signal
                peptidases I, score 51.12"
misc_feature    12869. .12910
                /gene="sip1"
                /note="PS00761 Signal peptidases I signature 3"
gene            13122. .14462
                /gene="sip2"
CDS             13122. .14462
                /gene="sip2"
                /note="SC2E1.14, sip2, probable signal peptidase I, len:
                446 aa; similar to many e.g. LEP_HAEIN signal peptidase I
                (EC 3.4.21.89) (349 aa), fasta scores; opt: 151, z-score:
                317.3 E(): 1.9e-10, 30.1% identity in 286 aa overlap.
                Contains PS00761 Signal peptidases I signature 3, Pfam
                match to entry signal_pept_I PF00461, Signal peptidases I,
                score 42.41 and probable transmembrane domain. Also
                similar to surrounding genes sip1 (E(): 9.7e-24, 44.9%
                identity in 274 aa overlap), sip3 (E(): 1.8e-28, 56.3%
                identity in 245 aa overlap), and sip4 (E(): 1.1e-09, 35.9%
                identity in 223 aa overlap). Note that after aa 325 the
                ORF overlaps the downstream sip3 gene SC2E1.15, and
                positional base preference suggests that translation is in
                the other frame. No error in the cosmid sequence can be
                found to account for this"
                /codon_start=1
                /transl_table=11
                /product="putative signal peptidase I"
                /db_xref="PID:e1296634"
                /db_xref="PID:g3191990"
                /translation="MGSRGKPRGAPSSPAENLLPTGSRRTAAPSGGRSRAERRKLQRK
                VKRRRRRGAVKEIPLLVGVAVLIALVLKTFLVQAFVIPSGSMEQTIRIGDRVLVDKLT
                PWFGSEPQRGDVVVFRDPGGWLQGEQTTKKDDPVVVKQVKEGLAFIGLLPSDDEKDLI
                KRVVGVGGDHVKCCDKQGRVTVNGVPLTEDYLYPGDRPSRTPFDVTVPEGRLWVMGDH
                RSNSADSRAHQETDFGTVSQDEVVGRAMVIAWPFGHWTTLDEPKTYASVSDAASGSTA
                APELSHRVAPYDPNAMMELPTPAELPLVMGVVGLRRIRGRRRQRVRSWRGGCGGWRTV
                RSRRRGEPRTPRGEVRSGRGRRPRLRTRLRNRGRQGDERTARAGRTQRRDGGRGSRRD
                ATRDPARGQEAALLLEGAADPGRHRAGARPADQDVPGAGVLDPVLLDGEHPSDR"
misc_feature    13236. .13484
                /gene="sip2"
                /note="Pfam match to entry signal_pept_I PF00461, Signal
                peptidases I, score 42.41"
misc_feature    13755. .13796
                /gene="sip2"
                /note="PS00761 Signal peptidases I signature 3"
misc_feature    13859. .15213
                /note="97.8% match to EMBL:SLTK24SIP Y09474 Streptomyces
                lividans sip gene, from 5 to 1359"
RBS             14085. .14089
                /gene="sip2"
                /note="possible RBS upstream of sip3"
gene            14095. .15105
                /gene="sip3"
CDS             14095. .15105
                /gene="sip3"
                /note="SC2E1.15, sip3, probable signal peptidase I, len:
                336 aa; identical to, but longer than, S. lividans sip
                TR:E1228248 (EMBL:Y09474) signal peptidase I (291 aa),
                fasta scores; opt: 1997 z-score: 2287.4 E(): 0, 99.7%
                identity in 291 aa overlap, and similar to many others
                e.g. LEP_PHOLA signal peptidase I (EC 3.4.21.89) (203 aa),
                fasta scores; opt:222 z-score: 404.9 E(): 2.6e-15, 37.9%
                identity in 232 aa overlap. Contains PS00761 Signal
                peptidases I signature 3, 2x Pfam match to entry
                signal_pept_I PF00461, Signal peptidases I, scores 48.83
                and 46.79 and probable transmembrane domain. Also similar
                to surrounding genes sip1 (E(): 7.6e-20, 35.6% identity in
                264 aa overlap), sip2 (E(): 0, 56.3% identity in 245 aa
                overlap) and sip4 (E(): 6.6e-15, 36.3% identity in 237 aa
                overlap). Note that a more upstream start was chosen due
                to positional base preference"
                /codon_start=1
                /transl_table=11
                /product="putative signal peptidase I"
                /db_xref="PID:e1296635"
                /db_xref="PID:g3191991"
                /translation="MGDVAVGARSGHDGEENRGRPEERSGPAAVGAPDSGHGSGTEDG
                RVTNGQRGQDGHNGGTEDEGVGGTPPATPPAAKKQRSFWKELPILVGIALVLALLIKT
                FLVQAFSIPSSSMENTLQIGDRVLVDKLTPWFGSEPERGEVVVFHDPADWLAGEPTPD
                PNALQTVLSWIGLMPSAEEKDLIKRVIGVAGDTVECNKTGPLKVNGKALNEPYVYPGN
                TPCSDDDQGGRFKVTVPEGKIWVMGDHRQNSRDSRYNQSDKNGGMVPVDEVVGRAIVV
                AWPMNRWGTLPVPDTFDQDGLQARSSAAAALSVAPQGLAVAGVVPFVWWRRRRTAPAE
                TR"
misc_feature    14302. .14547
                /gene="sip3"
                /note="Pfam match to entry signal_pept_I PF00461, Signal
                peptidases I, score 48.83"
misc_feature    14632. .14907
                /gene="sip3"
                /note="Pfam match to entry signal_pept_I PF00461, Signal
                peptidases I, score 46.79"
misc_feature    14809. .14850
                /gene="sip3"
                /note="PS00761 Signal peptidases I signature 3"
RBS             15197. .15202
                /note="possible RBS upstream of sip4"
gene            15210. .15986
                /gene="sip4"
CDS             15210. .15986
                /gene="sip4"
                /note="SC2E1.16, sip4, probable signal peptidase I, len:
                258 aa; similar to many e.g. LEPS_BACSU signal peptidase I
                (EC 3.4.21.89) (184 aa), fasta scores; opt: 258 z-score:
                337.6 E(): 1.4e-11, 32.2% identity in 177 aa overlap.
```

Contains Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 49.57 and probable transmembrane domain. Also similar to upstream genes sip1 (E(): 8.8e-10, 36.5% identity in 219 aa overlap), sip2 (E(): 3.8e-13, 35.9% identity in 223 aa overlap) and sip3 (E(): 8.1e-15, 35.4% identity in 237 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296636"
/db_xref="PID:g3191992"
/translation="MGGESTTRTVPRGGGANKGPVGSRTGQRLSGIAVALGLVLFLGG
FAWGAVVYRPYTVPTSSMTPTIDAGDRVLAQRIDGADVRRGDVVVFKDATWANAPMVK
RVVAVGGDTVSCCQQGKLKVNGKVIDEPYLPAGTPAEISDFQTVTVPEGRLFLLGDER
RNSVDSTAHLTDAAAGTVSRGAVDARVDAVAWPMDGMLERPTGFRTLGDLSSPGPLRT
VVVLVIAGAVLILGGAAYGPLAKRAAASRARKGAEPAGGR"

misc_feature 15255. .15605
/gene="sip4"
/note="Pfam match to entry signal_pept_I PF00461, Signal peptidases I, score 49.57"

gene 15988. .16506
/gene="SC2E1.17"

CDS 15988. .16506
/gene="SC2E1.17"
/note="SC2E1.17, unknown, len: 172 aa; similar to C-terminus of M. tuberculosis hypothetical protein TR:O06205 (EMBL:Z95387) MTCY1A10.24 (351 aa), fasta scores; opt: 345 z-score: 375.2 E(): 1.2e-13, 44.7% identity in 161 aa overlap. Contains PS00893 mutT domain signature and Pfam match to entry mutT PF00293, Bacterial mutT protein, score 37.78"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.17"
/db_xref="PID:e1296637"
/db_xref="PID:g3191993"
/translation="MTPAGQDTYEGGLRRVARVVLLDPEDRILLLHGHEPDDPADDWW
FTPGGGVEGDETRAEAARRELLEETGITDVELGPVLWRRRCSFPFAGRRWDQDEWYYL
ARTARTATEAVGPGLTELERRSVAGARWWTCEELTRARETVYPTRLAELLTTLLDEGP
PAGPVTLDTEIV"

misc_feature 16069. .16200
/gene="SC2E1.17"
/note="Pfam match to entry mutT PF00293, Bacterial mutT protein, score 37.78"

misc_feature 16132. .16191
/gene="SC2E1.17"
/note="PS00893 mutT domain signature"

gene 16563. .16871
/gene="SC2E1.18"

CDS 16563. .16871
/gene="SC2E1.18"
/note="SC2E1.18, unknown, len: 102 aa; highly similar to M. tuberculosis hypothetical protein YX32_MYCTU MTCY274.32c (101 aa), fasta scores; opt: 515 z-score: 1085.2 E(): 0, 75.0% identity in 100 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.18"
/db_xref="PID:e1296638"
/db_xref="PID:g3191994"
/translation="MSAEDLEKYETEMELKLYREYRDVVGLFKYVIETERRFYLTNDY
EMQVHSVQGEVFFEVSMADAWVWDMYRPARFVKQVRVLTFKDVNIEELNKSDLELPGG
"

gene 16930. .17322
/gene="SC2E1.19"

CDS 16930. .17322
/gene="SC2E1.19"
/note="SC2E1.19, unknown, len: 130 aa; similar to M. tuberculosis hypothetical protein YX29_MYCTU MTCY274.29c (128 aa), fasta scores; opt: 253 z-score: 276.8 E(): 3.5e-08, 39.7% identity in 121 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.19"
/db_xref="PID:e1296639"
/db_xref="PID:g3191995"
/translation="MIVGAGGGADMNARGAMGRYGETLAARRLTGAGMTVLERNWRCG
RTGEIDIVARDGDVLVVCEVKTRRGGAFEHPMAAVTPDKAERLRRLAERWIQTHGGAP
PGGVRIDLVGVLLPQRGAPVVEHARGVA"

gene 17322. .18947
/gene="SC2E1.20"

CDS 17322. .18947
/gene="SC2E1.20"
/note="SC2E1.20, unknown ATP/GTP binding protein, len: 541aa; similar to hypothetical proteins from many organisms e.g. YIFB_ECOLI (516 aa), fasta scores; opt: 356 z-score: 935.9 E(): 0, 37.0% identity in 513 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.20"
/db_xref="PID:e1296640"
/db_xref="PID:g3191996"
/translation="MGFARTCSVALVGVEGVVVEVQADLEPGVAAFTLVGLPDKSLTE
SRDRVRAAVVNSGAQWPQKKLTVGLSPASVPKSGSGFDLAVAASVLGAAERIDPRVLA
DIVMIGELGLDGRVRPVRGILPAVLAAAEAGYEQVVVPECAAPEASLVPGVSVLGVRS
LRQLIAVLADEPVPEEEQDLQGRPDPLMAGLRMPGTGAATGMHGMGAAQHDHGHDLAD
VVGQESARTAVEVAAAGRHHLFLEGPPGAGKTMLAERLPAILPRLCRAESLEVTAVHS
VAGLLPPGKPLIDVAPYCAPHHSATMQALVGGGPGIARPGAVSLAHRGVLFLDETPEF
SSHALDALRQPLEAGHVVIARSAGVVRFPARFLMVLAANPCPCGRFSRTDELCECPSS
AIRRYQARLSGPLLDRVDLRVEVERVTRSQLAGDGPRGDSTRAVADRVRAARDRAAAR
LAGTPWRSNSEVPGRELRTRWRAAAGALDEAERSLERGVLTARGLDRVLRVAWTVADL
VGHDRPDARDVALALQLRTGVPRGVPMALGALA"

misc_feature 18048. .18071
/gene="SC2E1.20"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

gene 18944. .20092
/gene="SC2E1.21"

CDS 18944. .20092
/gene="SC2E1.21"
/note="SC2E1.21, unknown, len: 382 aa; N-terminus is Gly-rich, C-terminus is similar to the N-terminus of several hypothetical proteins e.g. SMF_ECOLI SMF protein (374 aa), fasta scores; opt: 402 z-score: 277.5 E(): 3.2e-08, 40.7% identity in 194 aa overlap. This ORF appears to be interrupted by the following prophage at around aa 360, and to continue as the last part of SC2E1.37. The similarity to the M. tuberculosis hypothetical protein YX27_MYCTU MTCY274.27c (389 aa) (fasta scores; opt: 603 z-score: 511.5 E(): 3e-21, 46.8% identity in 267 aa overlap) stops at that position and restarts in SC2E1.37. Alternative start site at aa 118 sugested by positional base preference"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.21"
/db_xref="PID:e1296641"
/db_xref="PID:g3191997"
/translation="MSGNTRPPEGAAPSGDTGPGTSGTLFPEAAVSFGGGSPAEGGAM
TPPGAGDGPDGGEARWAGPGPGDGDGAVTGAGVAARGRSAAGTCGTGGGGAGSGARGG
GPEDRELLGRVFLARVFEPGDEAGGRWVRERGAPEVVRRLREGGRALPGVSEKRWAGL
CARAGRADPGRDLAVARSAGARFVVPGTAEWPGQLDDLGDARPLGLWVRGGPSLRMWA
LRSVAVVGARACTEYGAHMAATLAAGLAERGWVVVSGGAYGIDGAAHRGALGAGGATA
AVLACGVDRPYPPGHTALITRIAEQGLVVGELPPGDHPTPSRFILRNRVIAALTRGTV
VVEAAHRSGSLVTARAARRLGRHVMGVPTPLSGGAALAPIRFRLATAT"

misc_feature 20021. .32836
/note="probable prophage: has low GC (approx 66-68%), and some ORF's with similarities to bacteriophage ORF's. The ORF SC2E1.21 appears to have been interrupted by this prophage and apparently continues as the last part of SC2E1.37"

gene 20271. .20582
/gene="SC2E1.22"

CDS 20271. .20582
/gene="SC2E1.22"
/note="SC2E1.22, unknown prophage gene, len: 103 aa"

```
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.22"
                /db_xref="PID:e1296642"
                /db_xref="PID:g3191998"
                /translation="MTTQNSNHPCACGSYAFEVLIHENVGGDKVWQQKTTGCAATTQS
                TFAPGHDAKLKSLLIAAGVGGHPVRQTTRDTVVVKDALKVAADLGWRDLVGEAIAKGS
                S"
gene            complement(20579. .20869)
                /gene="SC2E1.23c"
CDS             complement(20579. .20869)
                /gene="SC2E1.23c"
                /note="SC2E1.23c, unknown prophage gene, len: 96 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.23c"
                /db_xref="PID:e1296643"
                /db_xref="PID:g3191999"
                /translation="MTTRYPALPQSYAITRDRDTVIPRYLVVTALRGGQMAGLSIRVY
                KIVGETGETEEVSAREWSVDQMPNLPLIDTRWPLCLCWKCNGNGTKGELRRR"
RBS             20912. .20916
                /note="possible RBS upstream of SC2E1.23"
misc_feature    20920. .21099
                /gene="SC2E1.24"
                /note="Pfam match to entry gntR PF00392, Bacterial
                regulatory proteins, gntR family, score 43.92"
gene            20920. .21114
                /gene="SC2E1.24"
CDS             20920. .21114
                /gene="SC2E1.24"
                /note="SC2E1.24, possible prophage transcriptional
                regulator, len:64 aa; similar to the N-terminus of several
                regulatory proteins e.g. Streptomyces phaeochromogenes
                plasmid pJV1 TR:Q54677 (EMBL:U23762) traR (245 aa), fasta
                scores; opt:137 z-score: 268.5 E(): 1e-07, 39.3% identity
                in 61 aa overlap. Contains Pfam match to entry gntR
                PF00392, Bacterial regulatory proteins, gntR family, score
                43.92"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /db_xref="PID:e1296644"
                /db_xref="PID:g3192000"
                /translation="MAEIISKRIADGTYPPGTRVPGIVELSAEFGIAASTAQKALAHL
                RGTGEVRTELGLGTFVADAK"
RBS             21200. .21204
                /note="possible RBS upstream of SC2E1.25"
gene            21209. .21910
                /gene="SC2E1.25"
CDS             21209. .21910
                /gene="SC2E1.25"
                /note="SC2E1.25, unknown prophage gene, len: 233 aa"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.25"
                /db_xref="PID:e1296645"
                /db_xref="PID:g3192001"
                /translation="MATTDGDEPPEEREFLTGDTPFVRVREALERLAALNEGPRVEQR
                VDRVDQIQNVFIGQGQDTHGREVAERAATRLAEQRQKFHFDFLNHALKQAEWTFRLSV
                WFMTGGALVILAGAVLALVHAGNPDLSYLPLVTSLTGALITVGGGALALHSKRAMANL
                TKAAEDNENKIDIDHKLEVATTFIDRVQDSQAKDRLNSAAALKALGMDAQPETMVNRL
                LPDPQTKEIDPGEPT"
gene            complement(21969. .22166)
                /gene="SC2E1.26c"
CDS             complement(21969. .22166)
                /gene="SC2E1.26c"
                /note="SC2E1.26c, unknown prophage gene, len: 65 aa; very
                hydrophobic"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.26c"
                /db_xref="PID:e1296646"
                /db_xref="PID:g3192002"
                /translation="MNLPLGVKVLIVVICTLISVIVAMVAGFLSHASGSSVGQAILYG
                GGAFAGSLLLCLAVLSALGVL"
RBS             complement(22177. .22180)
                /note="possible RBS upstream of SC2E1.26c"
gene            complement(22505. .22867)
                /gene="SC2E1.27c"
CDS             complement(22505. .22867)
                /gene="SC2E1.27c"
                /note="SC2E1.27c, unknown prophage gene, len: 120 aa;
                contains hydrophobic region near centre"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein SC2E1.27c"
                /db_xref="PID:e1296647"
                /db_xref="PID:g3192003"
                /translation="MSEIPAIRRRRASQFDHIDCGRADECPVHEPLLGHPTRKVPSQL
                DGAMTLAILAVAGVVSIFLFALKGVLDQVPDVIDSASRARDAWQRFKRKANEPPSPTV
                LRELPPPAEDEQEPPAAA"
gene            23631. .23846
                /gene="SC2E1.28"
CDS             23631. .23846
                /gene="SC2E1.28"
                /note="SC2E1.28, possible prophage transcriptional
                regulator, len: 71 aa; similar to the N-terminus of many
                e.g. TR:Q37946 Salmonella typhimurium bacteriophage L
                repressor protein C2 (220 aa), fasta scores; opt: 127
                z-score: 227.1 E():2e-05, 45.7% identity in 46 aa overlap.
                Contains probable helix-turn-helix motif at aa 21 to 42
                (Score 1829, +5.42 SD)"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /db_xref="PID:e1296648"
                /db_xref="PID:g3192004"
                /translation="MPARTFSPDALRQIRRQRGVTQYKLAASLGVHNTAISHYENGHR
                RPDVETLATIADVLGARMDDFVTARAA"
RBS             23996. .24001
                /note="possible RBS upstream of SC2E1.29"
gene            24008. .26278
                /gene="SC2E1.29"
CDS             24008. .26278
                /gene="SC2E1.29"
                /note="SC2E1.29, unknown prophage ATP binding protein,
                len: 756 aa; some similarity to TR:Q38030 (EMBL:X76288)
                Streptomyces Bacteriophage phi-C31 early region ORF 9 (519
                aa), fasta scores; opt: 360 z-score: 374.4 E(): 1.3e-13,
                24.4% identity in 553 aa overlap, and to a putative
                prophage ORF from M. tuberculosis TR:O06608 (EMBL:Z95586)
                MTCY336.22 (471 aa), fasta scores; opt: 655 z-score: 535.5
                E(): 1.4e-22, 29.8% identity in 416 aa overlap. Contains
                2x PS00017 ATP/GTP-binding site motif A (P-loop)"
                /codon_start=1
                /transl_table=11
                /product="putative ATP binding protein"
                /db_xref="PID:e1296649"
                /db_xref="PID:g3192005"
                /translation="MSYTTSEARDLAATQSGLDVPEHDTDRAAYFARLGFGVFPMPTN
                KKMPRAKGWPDLAMTTVEEVEEMWRSFTPSGWQNIALTLDGWTVVDVDPKNGGTASLE
                KLTAEYELPVTRTHHTASGPDSLHFIYRADPDRPIKSGPLSSRYPGIDIKTGRGSLIV
                APGSVINGRRYEVVSAAEPAVAPSWLSEIRGDAQYPVVPSARSRTPSRASAGHSRAAT
                LAELRALPPDDPGRGNGWLTQVAGHLARNHTDHDSYLHELRAIDSESEVPHDKDRFMK
                TADSIWEREQAKERPKAAAPDKPKSILWHHEHTDLGNGRRLRDLHGEDMRHAPGWGWL
                TWDGRRWARGDGQARRFMHSVADAISEAIESMEEKQRSDDLMRWFGRSCSGPGISAAL
                NEASVMLPIETEISDFDADPHKLLVGNGVVDLRTGELLPVDRKYLLTRGTTVEYDPNA
                DCPMWMDFLGWAFQGDIEMIEYIQRMFGMCLIGNNAHQVAFFLYGPGRSGKTTLTRVL
                SRLLGDYATSADLSVFNESSSGHNEPLARLAGARLVVFSETRQGQRINEAQFKKFTGE
                DTLTASYKNKSWFEFLPAFTPVMFGNAQPSIAFDSGVERRMKVIPMRAQISDGQKNPK
                LVEQMMLNEGPAIMAWAVEGARLTAAESFVPDPPQVAQAVKEYKRENDHIGDFIEECL
                VFDEGATVASDVVWGRYSKWIEAGGVDFVKKADDRAGKSLLVRMLKTWSIENGTPIES
                FKSSNTRKLRGVALND"
misc_feature    25490. .25513
                /gene="SC2E1.29"
                /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
```

misc_feature 26147. .26170
/gene="SC2E1.29"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
RBS 26551. .26554
/note="possible RBS upstream of SC2E1.30"
gene 26562. .27113
/gene="SC2E1.30"
CDS 26562. .27113
/gene="SC2E1.30"
/note="SC2E1.30, unknown prophage gene, len: 183 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.30"
/db_xref="PID:e1296650"
/db_xref="PID:g3192006"
/translation="MNDLDRRGGRIFGRVDGEVDWPSGVDSPYRLMKLRDPEKYEREV
ARQVDAQRVIVEWAEWYGLKRAKMACCPRWLTHKVSRRCNSRCTAFGGGATDHRWLEH
GVGWVLDGQPVALTSASYGISDEYRARFEWWVNEDPRLAVALGGQGWYGYGSTQVLVW
RTDVIEVMEPAPESVRALGYPQY"
gene 27578. .27769
/gene="SC2E1.31"
CDS 27578. .27769
/gene="SC2E1.31"
/note="SC2E1.31, possible prophage transcriptional
regulator, len: 63 aa; similar to the N-terminus of many
e.g. TR:E264367 (EMBL:X87420) Salmonella typhimurium
bacteriophage ES18 gene C2 (216 aa), fasta scores; opt:
133 z-score: 207.9E(): 0.00024, 35.8% identity in 53 aa
overlap. Contains probable helix-turn-helix motif at aa 17
to 38 (Score 1786, +5.27 SD)"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="PID:e1296651"
/db_xref="PID:g3192007"
/translation="MFSGRKLKEARNKRGMSQAAVAARMGLSGLYVSYWENGRYAPNG
VNMVLLLRALGCELEEVTE"
RBS 27754. .27759
/gene="SC2E1.31"
/note="possible RBS upstream of SC2E1.32"
gene 27766. .28116
/gene="SC2E1.32"
CDS 27766. .28116
/gene="SC2E1.32"
/note="SC2E1.32, unknown prophage gene, len: 116 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.32"
/db_xref="PID:e1296652"
/db_xref="PID:g3192008"
/translation="MNAPETEFADAGWVAITRDEGGRVGVKPVRMFELTGEGVTALTK
NADGLMVPDPHAVEIINTDPLHAAKLAWLHKLVGVAERCTTDEARADLRRIAEWLIDW
EPGDPGLRLADAEA"
gene 28184. .28672
/gene="SC2E1.33"
CDS 28184. .28672
/gene="SC2E1.33"
/note="SC2E1.33, unknown prophage gene, len: 162 aa;
contains probable helix-turn-helix motif at aa 11 to 32
(Score 1192, +3.25 SD)"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.33"
/db_xref="PID:e1296653"
/db_xref="PID:g3192009"
/translation="MVVVLAQQGIGRNEISRRTGISTASVSRIAAAHGIEFDGSQTEV
ALKARIAELKARQAGIALGLVEDITSARGLLRMATTHRDFAFAAKAISDLTQSAQRMT
PEITNEEELEEAKEGLRDLYFQMGLMREGFEAKYGVPLDSDEGRELWDKARQEDMENE
QP"
RBS 28644. .28649
/gene="SC2E1.33"
/note="possible RBS upstream of SC2E1.34"
gene 28659. .29075
/gene="SC2E1.34"
CDS 28659. .29075
/gene="SC2E1.34"
/note="SC2E1.34, unknown prophage gene, len: 138 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.34"
/db_xref="PID:e1296654"
/db_xref="PID:g3192010"
/translation="MSNRDFSEDVQGPRSTDGPSVDDLRAEMYGGQVAGRFHYRESRT
ERDSAVIQGTPVLGDGTMTDDQAAEIRKSFGIGADLRHAAQAAESYMDRRDPTSGHTI
RTYGNGRVVDLDAEPTQTGSDRVEYPRVWYGDDPGA"
RBS 29481. .29484
/note="possible RBS upstream of SC2E1.35"
gene 29493. .29936
/gene="SC2E1.35"
CDS 29493. .29936
/gene="SC2E1.35"
/note="SC2E1.35, unknown prophage gene, len: 147 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.35"
/db_xref="PID:e1296655"
/db_xref="PID:g3192011"
/translation="MAATKVFRPEFKTRDGRKVSFGANYTHDTKAKRRAGLTTAGRPK
VETVGAGTVVRSPRRWLPWAVFGTNSAVLGGTWAVAELGPAVGGLVVGVALWAGYRIA
RWTAGEWAYFATKSRQIRGGRQVAEFASEGLPPQAHIPGEKVTRP"
RBS 30237. .30240
/note="possible RBS upstream of SC2E1.36"
gene 30248. .30754
/gene="SC2E1.36"
CDS 30248. .30754
/gene="SC2E1.36"
/note="SC2E1.36, unknown prophage gene, len: 168 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.36"
/db_xref="PID:e1296656"
/db_xref="PID:g3192012"
/translation="MAAKKTETAEATPCECSMYDALPADLTEEQVASGDFEVLTTGCT
ATTKRQFAPGHDAKLKSALIRWGALGLEIRRNEGGVATSASPAKHAARYAFAHMVTAG
VKRAEAKAAEKAERAAARAAKKAAPAPEVIKAKVGRVTYQGRMDGDHFVYEVKGQERR
TLKFQPAA"
gene 30824. .33214
/gene="SC2E1.37"
CDS 30824. .33214
/gene="SC2E1.37"
/note="SC2E1.37, unknown prophage gene, len: 796 aa; shows
very limited similarity to Rhodococcus opacus TR:G2935035
(EMBL:AF003948) putative transposase homolog (242 aa),
fasta scores; opt: 215 z-score: 280.3 E(): 2.2e-08, 33.3%
identity in 105 aa overlap. Contains PS00397 Site-specific
recombinases active site and Pfam match to entry
recombinase PF00239, Site-specific recombinases, score
36.16. The last part of this ORF (aa 672 onwards) appears
to be a continuation of SC2E1.21, which was interrupted by
the prophage; the similarities to YX27_MYCTU MTCY274.27c
(389 aa) (fasta scores; opt: 246 z-score: 225.9 E():
2.4e-05, 38.8% identity in 116 aa overlap) continue after
aa 672. Contains probable coiled-coil from aa 449 to 505"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.37"
/db_xref="PID:e1296657"
/db_xref="PID:g3192013"
/translation="MYQRNTDALDVAMGVVERAVQPVRLRAVAYARVSTEEQAKGYGV
QAALKKILRYIGRKDWDHVGTYTDEGISGSLEAADREDLKRLMADAHRQPKPFDIVVV
SEGRAIGRVGRAFWRWVWALEDIGIYVGVVEDDYDNSTPEGRKKMRRDADYAETEWET
IRKRTQGGLQEKAEDGGWPGGQPPYGYEIACQGQKGKSHLVQCAAEVKTLRMAWAMVV
EEKLNTREVAARLNALNRFTRSGVPWSHSNLRAKLLSESTINARVIFRNPNRAHAGHG
AKFGEDGKPLHGETVIIKLKPIFEPGEVAALQVTLAQTSKGAGRPKPKPYPLSKRLFN
EHTGCNSHHVGMARTSREGRWYRCTGLAAKYPGDPTCTCKMVDADAMESAVWGEVVSL
LGDPDRLRAMAAEWVGMAAGDQVQHADRIADFDKQIANLDRAIASMATECVKAGLPAA

```
                     AIAQASAALLEERRQLADLRDEAAAWLEETEAAEQRARDLEALATAARTRLADMDPAE
                     QGEVLALLDVKVTITGPVPKPKLGLTCSLAEWFKVAGRLVPSELTDDMWAAAEPVVKA
                     WEPSNHKLRPGRLMLDAMFYKARTGCRWDDLPERFGPWKGIHSRYKTWRNCGVWDEIM
                     AALPVDGPGYRPVPEINLVPPFRVEGRVDPRVLAGPDIQEEVVRPETGVPGPATSALS
                     AGVHEMLRGDAVLVGDAAEVVELVGAMGELPPERRGPVLPTDLLESRTRQVLAGLPGR
                     GTATAGEVALRAQTTQDDAVARLYELRALGYVERHGDGWKLTRQAMISVRGGRSPC"
misc_feature         30905. .31387
                     /gene="SC2E1.37"
                     /note="Pfam match to entry recombinase PF00239,
                     Site-specific recombinases, score 36.16"
misc_feature         30911. .30937
                     /gene="SC2E1.37"
                     /note="PS00397 Site-specific recombinases active site"
gene                 33497. .34339
                     /gene="whiG"
CDS                  33497. .34339
                     /gene="whiG"
                     /note="SC2E1.38, whiG, RNA polymerase sigma factor WhiG,
                     len: 280 aa; identical to RPSW_STRCO RNA polymerase sigma
                     factor whiG (280 aa). Also highly similar toRPSD_BACSU RNA
                     polymerase sigma-D factor (sigma-28) (254 aa)(E(): 0,
                     43.0% identity in 242 aa overlap) and FLIA_ECOLI RNA
                     polymerase sigma factor for flagellar operon (sigma F)
                     (239 aa)(E(): 0, 42.8% identity in 229 aa overlap).
                     Contains PS00715 and PS00716 Sigma-70 factors family
                     signatures 1 and 2, Pfam match to entry sigma70 PF00140,
                     Sigma-70 factors, score 342.03 and probable
                     helix-turn-helix motif at aa 245 to 266 (Score 1657, +4.83
                     SD)"
                     /codon_start=1
                     /transl_table=11
                     /product="RNA polymerase sigma factor WhiG"
                     /db_xref="PID:e1296658"
                     /db_xref="PID:g3192014"
                     /translation="MPQHTSGSDRAAIPPAARDGGSVRPPAPSTLDELWRSYKTTGDE
                     RLREQLILHYSPLVKYVAGRVSVGLPPNVEQADFVSSGVFGLIDAIEKFDVDREIKFE
                     TYAITRIRGAMIDELRALDWIPRSVRQKARNVERAYATLEARLRRTPSESEVAVEMGI
                     AVEDLHAVFSQLSLANVVALEELLHAGGEGGGRLSLMDTLEDTAADNPVEVAEDRELR
                     RLLARAINTLPEREKTVVTLYYYEGLTLAEIGNVLGVTESRVSQIHTKSVLQLRAKLA
                     GFGR"
misc_feature         33626. .34315
                     /gene="whiG"
                     /note="Pfam match to entry sigma70 PF00140, Sigma-70
                     factors, score 342.03"
misc_feature         33725. .33766
                     /gene="whiG"
                     /note="PS00715 Sigma-70 factors family signature 1"
misc_feature         34229. .34309
                     /gene="whiG"
                     /note="PS00716 Sigma-70 factors family signature 2"
RBS                  34449. .34453
                     /note="possible RBS upstream of SC2E1.39"
gene                 34468. .35025
                     /gene="SC2E1.39"
CDS                  34468. .35025
                     /gene="SC2E1.39"
                     /note="SC2E1.39, possible transcriptional regulator, len:
                     185 aa; equivalent to Streptomyces roseosporus TR:P72569
                     (EMBL:U58281) tetR homologue whiR (179 aa) (86.6% identity
                     in 179 aa overlap) and similar to several putative
                     transcriptional regulators of the tetR family e.g.
                     TR:E1247678 putative transcriptional regulator SC10A5.11
                     (S.coelicolor) (198 aa), fasta scores; opt: 434 z-score:
                     720.5 E(): 6.8e-33, 41.5% identity in 171 aa overlap.
                     Contains probable helix-turn-helix motif at aa 29 to 50
                     (Score 1618, +4.70 SD) and Pfam match to entry tetR
                     PF00440, Bacterial regulatory proteins, tetR family, score
                     29.18"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /db_xref="PID:e1296659"
                     /db_xref="PID:g3192015"
                     /translation="MAEHRSMQRAALLDAARSLLSDGGTEALTFPALAERTGLARSSV
                     YEYFRSRAAVVEELCAVDFPVWAAEVEAAMERETAPDAKVEAYVRTQLDLVGDRRHRA
                     VVAISASELDAGAREKIRAAHGGLIAMIVEALGEMGHGQPRLAAMLLQGVVDAAVRRI
                     ELGAAEEPGAITEAAVSMALRGVRG"
misc_feature         34501. .34641
                     /gene="SC2E1.39"
                     /note="Pfam match to entry tetR PF00440, Bacterial
                     regulatory proteins, tetR family, score 29.18"
gene                 complement(34979. .35659)
                     /gene="SC2E1.40c"
CDS                  complement(34979. .35659)
                     /gene="SC2E1.40c"
                     /note="SC2E1.40c, probable secreted protein, len: 226 aa;
                     some similarity to two putative secreted M. tuberculosis
                     proteins YX22_MYCTU MTCY274.22 (249 aa), fasta scores;
                     opt: 315 z-score: 268.8 E(): 9.7e-08, 40.7% identity in
                     145 aa overlap (also upstream of rpsB) and Y0GK_MYCTU
                     MTCY13D12.20 (407 aa), fasta scores; opt: 218 z-score:
                     176.1 E(): 0.014, 34.1% identity in 164 aa overlap.
                     Contains probable N-terminal signal sequence."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SC2E1.40c"
                     /db_xref="PID:e1296660"
                     /db_xref="PID:g3192016"
                     /translation="MRARARVVVLSAVLAVLAVPTAQAGRAAGPPPPATAAAAGWALP
                     YAVGPMGRVDPAVPPVARHWPVGTRPAVLRGWDPPANPYGPGHRGVDLAAGGGAPVRA
                     VAAGRVSFAGPVGGRGVVSVELTGTGEPPLRTTFAPVDAAVRKGDEVAAGEPLGTVQP
                     TGSHCTACLHWGLLRGEVYLNPLSLLPPWLLDAGPSRLLPVAGVPVPGRGGISPVRRA
                     TPWRPPPP"
misc_feature         complement(35024. .38962)
                     /note="overlap with cosmid 6A9 (unfinished)"
RBS                  complement(35666. .35669)
                     /note="possible RBS upstream of SC2E1.40c"
misc_feature         35757. .37919
                     /note="97.4% match to AF034101 AF034101 Streptomyces
                     coelicolor ribosomal protein S2 (rpsB) and elongation
                     factor Ts (tsf) genes from 1 to 2154"
RBS                  35934. .35939
                     /note="possible RBS upstream of rpsB"
gene                 35948. .36880
                     /gene="rpsB"
CDS                  35948. .36880
                     /gene="rpsB"
                     /note="SC2E1.41, rpsB, 30S ribosomal protein S2, len: 310
                     aa; almost identical to RS2_STRCO 30S ribosomal protein S2
                     (242 aa), E(): 0, 96.7% identity in 242 aa overlap and
                     also highly similar to many others e.g. RS2_ECOLI 30S
                     ribosomal protein S2 (240 aa), fasta scores; opt: 744
                     z-score: 950.7 E(): 0, 50.4% identity in 224 aa overlap.
                     Contains PS00962 Ribosomal protein S2 signature 1 and
                     PS00211 ABC transporters family signature (probably a
                     false positive). Also contains Pfam matches to entry S2
                     PF00318, Ribosomal protein S2, score 304.89 and to entry
                     60s_ribosomal PF00428, 60s Acidic ribosomal protein, score
                     24.36. Note that this sequence is longer than RS2_STRCO
                     due to a sequence difference around 235 aa. This sequence
                     has a C-terminal extension not present in other rpsB. The
                     M. tuberculosis protein also has a C-terminal extension
                     but with a differentsequence. Our sequence has been
                     checked, and is also confirmed by sequence from the
                     overlapping cosmid 6A9."
                     /codon_start=1
                     /transl_table=11
                     /product="30S ribosomal protein S2"
                     /db_xref="PID:e1296661"
                     /db_xref="PID:g3192017"
                     /translation="MAVVTMRELLESGVHFGHQTRRWNPKMKRFIFTERNGIYIIDLL
                     QSLSYIDRAYEFVKETVAHGGTVMFVGTKKQAQEAIAEQATRVGMPYVNQRWLGGMLT
                     NFSTVYKRLQRLKELEQIDFEDVAASGLTKKELLVLSREKAKLEKTLGGIREMSKVPS
                     AVWIVDTKKEHIAVGEARKLNIPVVAILDTNCDPDEVDYKIPGNDDAIRSVTLLTRVI
                     ADAVAEGLIARSGAAGGAKGDKAAGEPLAAWERDLLEGEKAEKKDDAEAAEKPAEAPA
                     AEAPAAEAAEAPAAEAAPAEEPAAEAPAADAEQA"
```

```
     misc_feature    35963. .35998
                     /gene="rpsB"
                     /note="PS00962 Ribosomal protein S2 signature 1"
     misc_feature    35972. .36625
                     /gene="rpsB"
                     /note="Pfam match to entry S2 PF00318, Ribosomal protein
                     S2, score 304.89"
     misc_feature    36461. .36505
                     /gene="rpsB"
                     /note="PS00211 ABC transporters family signature"
     misc_feature    36740. .36856
                     /gene="rpsB"
                     /note="Pfam match to entry 60s_ribosomal PF00428, 60s
                     Acidic ribosomal protein, score 24.36"
     gene            36975. .37811
                     /gene="tsf"
     CDS             36975. .37811
                     /gene="tsf"
                     /note="SC2E1.42, tsf, elongation factor Ts, len: 278.
                     Almost identical to EFTS_STRCO elongation factor TS
                     (EF-TS) (276 aa), fasta scores; opt: 1695 z-score: 2425.0
                     E(): 0, 97.8% identity in 278 aa overlap and highly
                     similar to many others e.g. EFTS_ECOLI (282 aa), fasta
                     scores; opt: 321 z-score: 681.2 E(): 1e-30, 38.4% identity
                     in 276 aa overlap. Contains PS01126 Elongation factor Ts
                     signature 1"
                     /codon_start=1
                     /transl_table=11
                     /product="elongation factor Ts"
                     /db_xref="PID:e1296662"
                     /db_xref="PID:g3192018"
                     /translation="MANYTAADVKKLRELTGAGMMDCKKALDEAEGNVEKAVEALRIK
                     GQKGVAKREGRSAENGAVVSIIADDNTSGVLVELKCETDFVAKGEKFQNVATAIAEHV
                     AKAAPADLEALLASEIEPGKTVQAYVDEANANLGEKIVLDRFAQFSGGFVTAYMHRTM
                     PDLPPQIGVLVELDKPNAEIAKGVAQHIAAFAPKYLSKEDVPAEVVESERRVAEETTR
                     AEGKPEAALPKIVEGRLNGFFKDATLLGQPYALDNKKSVQKVLEEAGVSLKRFSRIKV
                     GI"
     misc_feature    37008. .37055
                     /gene="tsf"
                     /note="PS01126 Elongation factor Ts signature 1"
     RBS             38004. .38007
                     /note="possible RBS upstream of pyrH"
     gene            38017. .38778
                     /gene="pyrH"
     CDS             38017. .38778
                     /gene="pyrH"
                     /note="SC2E1.43, pyrH, uridylate kinase, len: 253 aa;
                     highly similar to many e.g. PYRH_ECOLI uridylate kinase
                     (EC 2.7.4.-) (240 aa), fasta scores; opt: 668 z-score:
                     968.6 E(): 0, 45.9% identity in 233 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="uridylate kinase"
                     /db_xref="PID:e1296663"
                     /db_xref="PID:g3192019"
                     /translation="MSSTKAEKSDDGKVSGRFMLKLSGEAFSGGGGLGVDPDVVHAIA
                     REIAAVVRDGAQIAIVIGGGNFFRGAELQVRGMDRARSDYMGMLGTVMNCLALQDFLE
                     KEGVDCRVQTAITMGQVAEPYIPLRAVRHLEKGRVVIFGAGMGMPYFSTDTTAAQRAL
                     EIDAEALLMGKNGVDGVYDSDPKTNPDAVKFDALGYGEVITRDLKVADATAVTLCRDN
                     SLPIVVFELLKEGNIARAVKGEKIGTLVGDQGSRD"
BASE COUNT      6233 a  13696 c  13758 g   5275 t
ORIGIN


  Query Match             7.0%;  Score 38.2;  DB 36;  Length 38962;
  Best Local Similarity   54.7%;  Pred. No. 0.52;
  Matches   76;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy      46 GTGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAG 105
           | || ||| ||    | |  |  ||| || | | | |||   |    || | |    ||
Db   35657 GAGGGCGCGTGCGCGAGTGGTGGTGCTGTCGGCCGTCCTGGCCGTCCTGGCGGTCCCGAC 35598

Qy     106 TGCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTG 165
```

```
           || | || | ||| || ||  |    | |  | || |||| | ||| ||||| | |||
Db   35597 CGCGCAGGCCGGCCGGGCCGCGGGCCCGCCGCCCCCGGCGACGGCGGCGGCTGCGGGGTG 35538

Qy     166 TGCGCTGCAGGGCGGGGAG 184
           || ||||    || || |
Db   35537 GGCCCTGCCCTACGCGGTG 35519


RESULT   9
AC004840/c
LOCUS           AC004840   162485 bp    DNA                HTG       12-JUN-1998
DEFINITION      *** SEQUENCING IN PROGRESS *** Homo sapiens clone DJ0607J02; HTGS
                phase 1, 12 unordered pieces.
ACCESSION       AC004840
NID             g3213157
KEYWORDS        HTG; HTGS_PHASE1.
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
                Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 162485)
  AUTHORS       Waterston,R.H.
  TITLE         The sequence of Homo sapiens clone
  JOURNAL       Unpublished
REFERENCE       2  (bases 1 to 162485)
  AUTHORS       Waterston,R.H.
  TITLE         Direct Submission
  JOURNAL       Submitted (12-JUN-1998) Genome Sequencing Center, Washington
                University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                MO 63108, USA
COMMENT         ***                                                          ***
                *** WARNING: Phase 1 High Throughput Genome Sequence ***
                ***                                                          ***
                * This sequence is unfinished. It consists of 12 contigs for
                * which the order is not known; their order in this record is
                * arbitrary. In some cases, the exact lengths of the gaps
                * between the contigs are also unknown; these gaps are  presented
                * as runs of N as a convenience only. When sequencing is complete,
                * the sequence data presented in this record will be replaced
                *by a single finished sequence with the same accession number.
                *       1     1548: contig of 1548 bp in length
                *    1549     1566: gap of unknown length
                *    1567     3038: contig of 1472 bp in length
                *    3039     3056: gap of unknown length
                *    3057     4512: contig of 1456 bp in length
                *    4513     4530: gap of unknown length
                *    4531     5985: contig of 1455 bp in length
                *    5986     6003: gap of unknown length
                *    6004    11153: contig of 5150 bp in length
                *   11154    11171: gap of unknown length
                *   11172    17312: contig of 6141 bp in length
                *   17313    17330: gap of unknown length
                *   17331    28043: contig of 10713 bp in length
                *   28044    28061: gap of unknown length
                *   28062    41728: contig of 13667 bp in length
                *   41729    41746: gap of unknown length
                *   41747    64066: contig of 22320 bp in length
                *   64067    64084: gap of unknown length
                *   64085    97366: contig of 33282 bp in length
                *   97367    97384: gap of unknown length
                *   97385   128426: contig of 31042 bp in length
                *  128427   128444: gap of unknown length
                *  128445   162485: contig of 34041 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162485
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ0607J02"
BASE COUNT     36267 a  42581 c  44937 g  38502 t    198 others
ORIGIN


  Query Match             6.8%;  Score 37.4;  DB 18;  Length 162485;
```

```
  Best Local Similarity   53.0%;  Pred. No. 1.1;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    180 GGGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCT 239
          ||  |  ||| |   |    || ||||   || ||| |||| || ||| ||||||||||
Db  90953 GGTGGGCACTGGGACCGCTCCAATCTCCCACAGGGACGGCAGGATGTGAGGCCCGGGGCT 90894

Qy    240 GAGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGA 299
          |  |      ||  |  | ||||  |||| | |    |||   |||    ||    | |
Db  90893 GCAGCACTCACCACCCCACGGGGTTGATCTTGTACCAGAACTTTTTCTGTTCGATCATGT 90834

Qy    300 CTGTCCCTACGGCACCTTCGGGATGGATTGC 330
          |  | |   | |  |||| | |  ||  | |
Db  90833 CCTTGCACTCTGTCCCTTGGTGGGGGGATTC 90803


RESULT  10
MUSFISP12B
LOCUS       MUSFISP12B   2267 bp    mRNA            ROD       09-JUL-1991
DEFINITION  Mouse FISP-12 protein (fisp-12) mRNA, complete cds.
ACCESSION   M70642
NID         g193315
KEYWORDS    cysteine-rich protein; growth factor-inducible gene.
SOURCE      Mus musculus (sub_species domesticus) liver/kidney cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 2267)
  AUTHORS   Ryseck,R.-P., MacDonald-Bravo,H., Mattei,M.-G. and Bravo,R.
  TITLE     Structure, mapping and expression of fisp-12, a growth factor
            inducible gene encoding a secreted cysteine-rich protein
  JOURNAL   Cell Growth Differ. 2, 225-233 (1991)
  MEDLINE   91363290
FEATURES             Location/Qualifiers
     source          1. .2267
                     /organism="Mus musculus"
                     /sub_species="domesticus"
                     /db_xref="taxon:10090"
                     /tissue_type="liver/kidney"
                     /map="A3-B1"
     gene            138. .2230
                     /gene="fisp-12"
     CDS             138. .1184
                     /gene="fisp-12"
                     /codon_start=1
                     /product="FISP-12 protein"
                     /db_xref="PID:g193316"
                     /translation="MLASVAGPISLALVLLALCTRPATGQDCSAQCQCAAEAAPHCPA
                     GVSLVLDGCGCCRVCAKQLGELCTERDPCDPHKGLFCDFGSPANRKIGVCTAKDGAPC
                     VFGGSVYRSGESFQSSCKYQCTCLDGAVGCVPLCSMDVRLPSPDCPFPRRVKLPGKCC
                     KEWVCDEPKDRTAVGPALAAYRLEDTFGPDPTMMRANCLVQTTEWSACSKTCGMGIST
                     RVTNDNTFCRLEKQSRLCMVRPCEADLEENIKKGKKCIRTPKIAKPVKFELSGCTSVK
                     TYRAKFCGVCTDGRCCTPHRTTTLPVEFKCPDGEIMKKNMMFIKTCACHYNCPGDNDI
                     FESLYYRKMYGDMA"
     polyA_signal    2224. .2230
                     /gene="fisp-12"
BASE COUNT      570 a    569 c    582 g    546 t
ORIGIN      chromosome 10.


  Query Match             6.6%;  Score 36.4;  DB 13;  Length 2267;
  Best Local Similarity   66.7%;  Pred. No. 1.3;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy    114 CAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGC 173
          | ||||   | ||  |||    ||||| |||| ||| |||||||||| || |||||
Db    260 CTGCCCCGCCGGCGTGAGCCTGGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCAA 319

Qy    174 AGGGCGGGGAGAAACTTG 191
            || |||||||   ||
Db    320 GCAGCTGGGAGAACTGTG 337


RESULT  11
MUSTGFB
LOCUS       MUSTGFB      2330 bp    mRNA            ROD       01-NOV-1991
DEFINITION  Mouse mRNA sequence.
ACCESSION   M80263
NID         g201945
KEYWORDS    transforming growth factor-beta.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 2330)
  AUTHORS   Brunner,A., Chinn,J., Neubauer,M.G. and Purchio,A.F.
  TITLE     Identification of a gene family regulated by transforming growth
            factor-beta
  JOURNAL   DNA Cell Biol. 10, 293-300 (1991)
  MEDLINE   91229699
FEATURES             Location/Qualifiers
     source          1. .2330
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="AKR-2B"
     mRNA            1. .2330
     CDS             204. .1250
                     /codon_start=1
                     /db_xref="PID:g201946"
                     /translation="MLASVAGPISLALVLLALCTRPATGQDCSAQCQCAAEAAPHCPA
                     GVSLVLDGCGCCRVCAKQLGELCTERDPCDPHKGLFCDFGSPANRKIGVCTAKDGAPC
                     VFGGSVYRSGESFQSSCKYQCTCLDGAVGCVPLCSMDVRLPSPDCPFPRRVKLPGKCC
                     EEWVCDEPKDRTAVGPALAAYRLEDTFGPDPTMMRANCLVQTTEWSACSKTCGMGIST
                     RVTNDNTFCRLEKQSRLCMVRPCEADLEENIKKGKKCIRTPKIAKPVKFELSGCTSVK
                     TYRAKFCGVCTDGRCCTPHRTTTLPVEFKCPDGEIMKKNMMFIKTCACHYNCPGDNDI
                     FESLYYRKMYGDMA"
     sig_peptide     204. .278
     mat_peptide     279. .1247
     polyA_signal    2291. .2296
     polyA_site      2318
BASE COUNT      589 a    598 c    595 g    548 t
ORIGIN


  Query Match             6.6%;  Score 36.4;  DB 13;  Length 2330;
  Best Local Similarity   66.7%;  Pred. No. 1.3;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy    114 CAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGC 173
          | ||||   | ||  |||    ||||| |||| ||| |||||||||| || |||||
Db    326 CTGCCCCGCCGGCGTGAGCCTGGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCAA 385

Qy    174 AGGGCGGGGAGAAACTTG 191
            || |||||||   ||
Db    386 GCAGCTGGGAGAACTGTG 403


RESULT  12
MUSFISP12A
LOCUS       MUSFISP12A   4128 bp    DNA             ROD       09-JUL-1991
DEFINITION  Mouse FISP-12 protein (fisp-12) gene, complete cds.
ACCESSION   M70641
NID         g193313
KEYWORDS    cysteine-rich protein; growth factor-inducible gene.
SOURCE      Mus musculus (sub_species domesticus) liver/kidney DNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 4128)
  AUTHORS   Ryseck,R.-P., MacDonald-Bravo,H., Mattei,M.-G. and Bravo,R.
  TITLE     Structure, mapping and expression of fisp-12, a growth factor
            inducible gene encoding a secreted cysteine-rich protein
  JOURNAL   Cell Growth Differ. 2, 225-233 (1991)
```

```
  MEDLINE   91363290
FEATURES             Location/Qualifiers
     source          1. .4128
                     /organism="Mus musculus"
                     /sub_species="domesticus"
                     /db_xref="taxon:10090"
                     /tissue_type="liver/kidney"
                     /map="A3-B1"
     TATA_signal     798. .804
                     /gene="fisp-12"
     exon            833. .1114
                     /gene="fisp-12"
                     /number=1
     mRNA            join(833. .1114,1202. .1424,1675. .1926,2059. .2270,
                     2642. .4006)
                     /gene="fisp-12"
     gene            join(1052. .1114,1202. .1424,1675. .1926,2059. .2270,
                     2642. .2938)
                     /gene="fisp-12"
     CDS             join(1052. .1114,1202. .1424,1675. .1926,2059. .2270,
                     2642. .2938)
                     /gene="fisp-12"
                     /codon_start=1
                     /product="FISP-12 protein"
                     /db_xref="PID:g193314"
                     /translation="MLASVAGPISLALVLLALCTRPATGQDCSAQCQCAAEAAPHCPA
                     GVSLVLDGCGCCRVCAKQLGELCTERDPCDPHKGLFCDFGSPANRKIGVCTAKDGAPC
                     VFGGSVYRSGESFQSSCKYQCTCLDGAVGCVPLCSMDVRLPSPDCPFPRRVKLPGKCC
                     KEWVCDEPKDRTAVGPALAAYRLEDTFGPDPTMMRANCLVQTTEWSACSKTCGMGIST
                     RVTNDNTFCRLEKQSRLCMVRPCEADLEENIKKGKKCIRTPKIAKPVKFELSGCTSVK
                     TYRAKFCGVCTDGRCCTPHRTTTLPVEFKCPDGEIMKKNMMFIKTCACHYNCPGDNDI
                     FESLYYRKMYGDMA"
     intron          1115. .1201
                     /gene="fisp-12"
                     /number=1
     exon            1202. .1424
                     /gene="fisp-12"
                     /number=2
     intron          1425. .1674
                     /gene="fisp-12"
                     /number=2
     exon            1675. .1926
                     /gene="fisp-12"
                     /number=3
     intron          1927. .2058
                     /gene="fisp-12"
                     /number=3
     exon            2059. .2270
                     /gene="fisp-12"
                     /number=4
     intron          2271. .2641
                     /gene="fisp-12"
                     /number=4
     exon            2642. .4006
                     /gene="fisp-12"
                     /number=5
BASE COUNT     1061 a   1039 c    984 g   1044 t
ORIGIN      chromosome 10.
```

```
  Query Match              6.6%;  Score 36.4;  DB 13;  Length 4128;
  Best Local Similarity   66.7%;  Pred. No. 1.4;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy     114 CAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGC 173
           | ||||   | ||  |||    ||||| |||| ||| ||||||||||| || |||||
Db    1261 CTGCCCCGCCGGCGTGAGCCTGGTGCTGGACGGCTGCGGCTGCTGCCGCGTCTGCGCCAA 1320

Qy     174 AGGGCGGGGAGAAACTTG 191
             || |||||||   ||
Db    1321 GCAGCTGGGAGAACTGTG 1338
```

```
RESULT  13
CJU13063
LOCUS       CJU13063     1927 bp    mRNA            VRT       28-AUG-1994
DEFINITION  Coturnix japonica cellular proto-oncogene protein Nov (nov) mRNA,
            complete cds.
ACCESSION   U13063
NID         g532696
KEYWORDS    .
SOURCE      Coturnix japonica.
  ORGANISM  Coturnix japonica
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
            Phasianidae; Phasianinae; Coturnix.
REFERENCE   1  (bases 1 to 1927)
  AUTHORS   Weiskirchen,R. and Bister,K.
  TITLE     Suppression in transformed avian fibroblasts of the nov
            protooncogene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1927)
  AUTHORS   Bister,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1994) Bister K., University of Innsbruck,
            Institute of Biochemistry, Peter-Mayr-Strasse 1a, Innsbruck,
            Austria, A-6020
FEATURES             Location/Qualifiers
     source          1. .1927
                     /organism="Coturnix japonica"
                     /db_xref="taxon:9094"
                     /clone="w23o"
                     /clone_lib="QEF cDNA library of R. Weiskirchen and K.
                     Bister"
                     /cell_type="fibroblast"
                     /dev_stage="embryo"
     gene            14. .1075
                     /gene="nov"
     CDS             14. .1075
                     /gene="nov"
                     /codon_start=1
                     /product="cellular proto-oncogene protein Nov"
                     /db_xref="PID:g532697"
                     /translation="MEPGGGHSLPVLLLLLLLLLLRPSEVNGREAPCPRPCGGRCPAE
                     PPRCAPGVPAVLDGCGCCLVCARQRGESCSPLLPCDESGGLYCDRGPEDGGGTGICMV
                     LEGDNCVFDGMIYRNGETFQPSCKYQCTCRDGQIGCLPRCNLGLLLPGPDCPFPRKIE
                     VPGECCEKWVCEPRDEVLLGGFAMAAYRQEATLGIDVSDSSANCIEQTTEWSACSRSC
                     GMGFSTRVTNRNQQCEMVKQTRLCMMRPCENEEPSDKKGKKCIRTKKSMKAVRFEYKN
                     CTSVQTYKPRYCGLCNDGRCCTPHNTKTIQVEFRCPQGKFLKKPMMLINTCVCHGNCP
                     QSNNAFFQPLDPMSSEAKI"
BASE COUNT      473 a    450 c    489 g    515 t
ORIGIN
```

```
  Query Match              6.6%;  Score 36;  DB 5;  Length 1927;
  Best Local Similarity   64.3%;  Pred. No. 1.7;
  Matches   54;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy     113 GCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTG 172
           || || | |    |   |    ||||||| |||| ||| ||||| ||||  ||||||||
Db     153 GCTGCGCCCCGGGGGTGCCCGCAGTGCTGGACGGCTGCGGCTGTTGCCTGGTGTGCGCCC 212

Qy     173 CAGGGCGGGGAGAAACTTGCTACC 196
               ||| ||||| |  |||| ||
Db     213 GGCAGCGCGGAGAGAGCTGCTCCC 236
```

```
RESULT  14
A28405
LOCUS       A28405       1975 bp    DNA             PAT       23-JUN-1995
DEFINITION  Sequence I that hybridises under stringent conditions to g.gallus
            nov gene.
ACCESSION   A28405
NID         g1248684
KEYWORDS    .
SOURCE      unidentified.
```

```
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1975)
  AUTHORS   .
  TITLE     NUCLEOTIDE SEQUENCES HYBRIDIZABLE WITH THE NOV GENE OF CHICKENS
  JOURNAL   Patent: WO 9300430-A 1 07-JAN-1993;
FEATURES             Location/Qualifiers
     source          1. .1975
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     source          1. .1975
                     /organism="artificial sequence"
                     /db_xref="taxon:29278"
BASE COUNT      522 a    451 c    486 g    516 t
ORIGIN


  Query Match             6.5%;  Score 35.8;  DB 6;  Length 1975;
  Best Local Similarity   52.9%;  Pred. No. 1.9;
  Matches  101;  Conservative    0;  Mismatches   87;  Indels    3;  Gaps    1;

Qy      6 CGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAATAA 65
          |||| ||||||||  |  |||||||| | |   |   |  ||||  ||   ||
Db     53 CGTCCTGCTGCTGCTCCTGCTCCTCCTCCGGCCGTGCGAGGTGAGCGGGCGGGAGGCGGC 112

Qy     66 TTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCGCTG 125
           |   |   |  ||||       | |||   |  |   |||       || || | |   |
Db    113 GTGCCCCCGGCCCTGCGGCGGGCGCTG---CCCCGCGGAGCCGCCGCGCTGCGCCCCGGG 169

Qy    126 CAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGGAGA 185
          |    | ||||| |||| ||| ||||||||||  ||||||||      ||| || ||
Db    170 AGTGCCCGCCGTGCTGGACGGCTGCGGCTGCTGCCTGGTGTGCGCCCGGCAGCGCGGCGA 229

Qy    186 AACTTGCTACC 196
          |  |||| ||
Db    230 GAGCTGCTCCC 240


RESULT  15
A28444
LOCUS       A28444       1975 bp    DNA             PAT       23-JUN-1995
DEFINITION  nov mRNA sequence.
ACCESSION   A28444
NID         g1248709
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1975)
  AUTHORS   .
  TITLE     NUCLEOTIDE SEQUENCES HYBRIDIZABLE WITH THE NOV GENE OF CHICKENS
  JOURNAL   Patent: WO 9300430-A 40 07-JAN-1993;
FEATURES             Location/Qualifiers
     source          1. .1975
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     source          1. .1975
                     /organism="artificial sequence"
                     /db_xref="taxon:29278"
BASE COUNT      522 a    451 c    486 g    516 t
ORIGIN


  Query Match             6.5%;  Score 35.8;  DB 6;  Length 1975;
  Best Local Similarity   52.9%;  Pred. No. 1.9;
  Matches  101;  Conservative    0;  Mismatches   87;  Indels    3;  Gaps    1;

Qy      6 CGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGCAATAA 65
          |||| ||||||||  |  |||||||| | |   |   |  ||||  ||   ||
Db     53 CGTCCTGCTGCTGCTCCTGCTCCTCCTCCGGCCGTGCGAGGTGAGCGGGCGGGAGGCGGC 112

Qy     66 TTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCGCGCTG 125
           |   |   |  ||||       | |||   |  |   |||       || || | |   |
Db    113 GTGCCCCCGGCCCTGCGGCGGGCGCTG---CCCCGCGGAGCCGCCGCGCTGCGCCCCGGG 169

Qy    126 CAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGGGGAGA 185
          |    | ||||| |||| ||| ||||||||||  ||||||||      ||| || ||
Db    170 AGTGCCCGCCGTGCTGGACGGCTGCGGCTGCTGCCTGGTGTGCGCCCGGCAGCGCGGCGA 229

Qy    186 AACTTGCTACC 196
          |  |||| ||
Db    230 GAGCTGCTCCC 240


Search completed: May 3, 1999, 16:12:45
Job time: 4404 sec
```